Bayesian SAE using Complex Survey Data
Lecture 8A: Advanced Topics

Jon Wakefield

Departments of Statistics and Biostatistics
University of Washington

Outline

Motivation

Overview of Lecture

In this lecture, we will first describe a first time model for estimating area-level U5MR over time, using a discrete hazards model.

So far we have carried out spatial modeling using discrete spatial models, sometimes referred to as Markov Random Field (MRF) models.

In this lecture we will also describe continuous spatial models, that allow estimation at a finer scale.

At the moment I view these as an elegant way of inducing spatial dependence between areal units (avoiding the arbitrariness of the neighbors in an MRF model), but others are promoting these models as a way of producing pixel-level surfaces, and so I will provide a critique of this approach.

U5MR Estimation in Space and Time

Kenyan Demographic Health Surveys

We base analyses on three Kenya DHS from 2003, 2008 and 2014.

These DHS use stratified (urban/rural, 8 regions), two-stage cluster sampling (enumeration areas, and then households).

All women age 15 to 49 who slept in the household the night before were interviewed in each selected household and response rates were high (above 95% for households in all surveys); these women asked to give what is known as full birth history:

- Birth dates of all children.
- Death dates for children who died.

DHS provides sampling (design) weights, assigned to each individual in the dataset, along with (jittered) GPS coordinates of the clusters.

The aim is small area estimation, in particular the U5MR and total deaths at the county level.

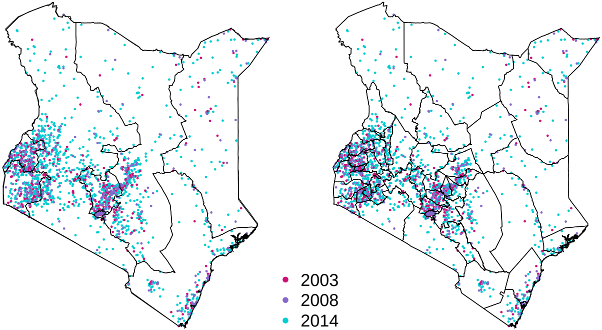


Figure 1: Cluster locations in the three Kenya DHS that we consider, with provincial (left) and Admin 1 (right) county boundaries.

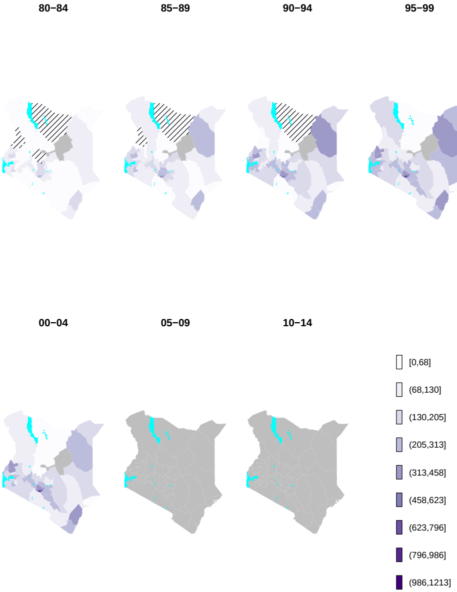


Figure 2: KDHS 2003: Number of births by Admin 1 area and 5-year period. Greyed out areas have no data hatched areas have less than 20 individual children.

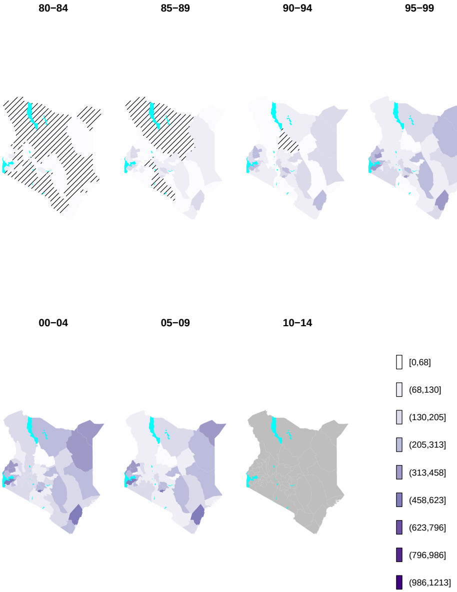


Figure 3: KDHS 2008: Number of births by Admin 1 area and 5-year period. Greyed out areas have no data hatched areas have less than 20 individual children.

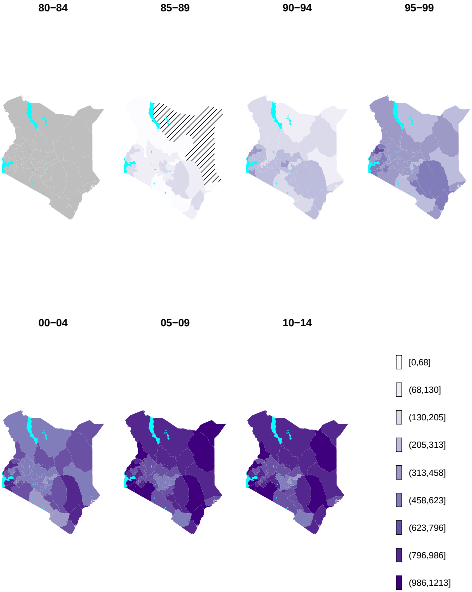


Figure 4: KDHS 2014: Number of births by Admin 1 area and 5-year period. Greyed out areas have no data hatched areas have less than 20 individual children.

Modeling Strategy

We will first describe the discrete hazard model which leads to an estimator of U5MR for a particular area i and time period t; call this estimator y_{it}, with design-based variance $\widehat{V}_{\text{DES},it}$.

Hierarchical Model:

1. The Data Model:

$$\underbrace{y_{it} \mid \lambda_{it} \sim \mathrm{N}\left(\lambda_{it}, \widehat{V}_{\text{DES},it}\right)}_{\text{Survey design acknowledged here}} \qquad .$$

2. The Space-Time (Random Effects) Prior:

$$\underbrace{\lambda_{it} = f(\text{ space } i, \text{ time } t)}_{\text{Smoothing here}}.$$

Discrete Hazards Model

As in Mercer et al. (2015) we assume a discrete hazard model, with six hazards for each of the age (monthly) bands: [0,1), [1,12), [12,24), [24,36), [36,48), [48,60].

For a generic period, area and survey:

$$
\begin{array}{lcl}
\text{Survival to 60 months} & = & \text{Survival in month 1} \\
 & \times & \text{Survival in month 2} \mid \text{survived to end of month 1} \\
 & \cdots & \\
 & \times & \text{Survival in month 60} \mid \text{survived to end of month 59.}
\end{array}
$$

In demography speak, and now for area i, period t and survey s:

$$
\begin{aligned}
1 - {}_{60}q_{0,its} &= \prod_{m=0}^{59} (1 - {}_1q_{m,its}) \\
&= (1 - {}_1q_{0,its}) \times (1 - {}_1q_{1,its}) \times (1 - {}_1q_{2,its}) \times \cdots \times (1 - {}_1q_{59,its}).
\end{aligned}
$$

Discrete Hazards Model

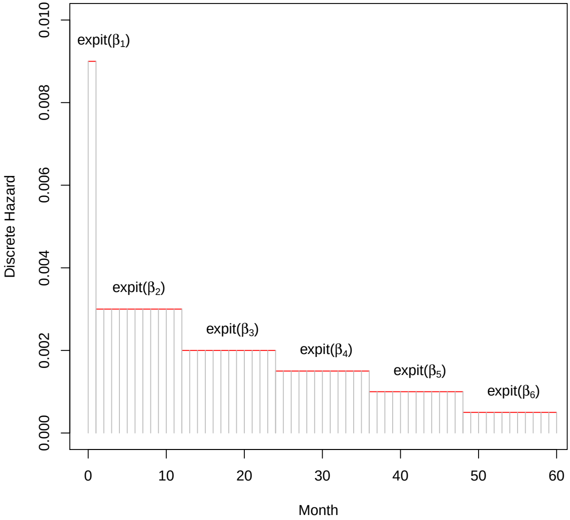
We calculate,

$$\begin{aligned}
\text{U5MR}_{its} &= {}_{60}q_{0,its} \\
&= 1 - \prod_{m=0}^{59} (1 - {}_1q_{m,its}) \\
&= 1 - \underbrace{(1 - {}_1q_{0,its}) \times (1 - {}_1q_{1,its}) \times (1 - {}_1q_{2,its}) \times \cdots \times (1 - {}_1q_{59,its})}_{\text{60 terms}} \\
&= 1 - \underbrace{\left[\frac{1}{1+\exp(\beta_{1,its})}\right] \times \left[\frac{1}{1+\exp(\beta_{2,its})}\right]^{11} \times \cdots \times \left[\frac{1}{1+\exp(\beta_{6,its})}\right]^{12}}_{\text{1+11+12+12+12+12 = 60 terms}}
\end{aligned}$$

Bottom line:

- For more complex designs we use weighted logistic regression (Binder, 1983) and obtain the hazards as the ratio of weighted deaths to weighted at risk in each month, with a standard error based on the design.

Discrete Hazards Model



"Meta-Analysis" Estimator

Combine survey information from S_t surveys in area i, period t:

$$
{}_{60}\widehat{q}_{0,it} = \text{expit}\left(\sum_{s=1}^{S_t} \underbrace{\left[\frac{\widehat{V}^{-1}_{\text{DES},its}}{\sum_{s=1}^{S_t} \widehat{V}^{-1}_{\text{DES},its}} \right]}_{\substack{\text{Weight for survey } s \text{ is} \\ \text{proportional to precision} \\ \text{of the survey}}} \text{logit}({}_{60}\widehat{q}_{0,its}) \right). \tag{1}
$$

This is the same estimator as the fixed-effects estimator used in meta-analysis.

Associated design-based variance (assuming independence of surveys):

$$
\widehat{V}_{\text{DES},it} = \left(\sum_{s=1}^{S_t} \widehat{V}^{-1}_{\text{DES},its} \right)^{-1},
$$

or, more informatively,

Precision of summary $=$ Sum of precisions of constituent surveys.

HIV epidemics result in selection bias

Let ${}_5q_{0\,l,k}(t)$ represent the true U5MR and ${}_5q_{0\,l,k}^{\star}(t)$ the biased (unadjusted for HIV) U5MR in survey k, province l and year t.

Walker et al. (2012) describe a method to provide an estimate of,

$$\mathrm{BIAS}_{l,k}(t) = \frac{{}_5q_{0\,l,k}^{\star}(t)}{{}_5q_{0\,l,k}(t)} \leq 1. \tag{2}$$

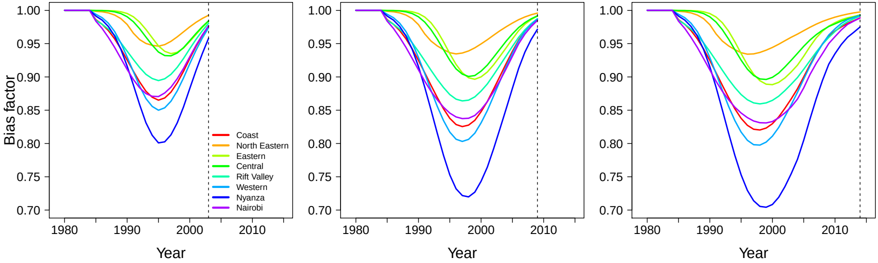


Figure 5: HIV adjustment ratios of reported U5MRs to “true” U5MRs, that is (2), by survey, over time (left is 2003, middle is 2008–2009, right is 2014), and in eight provinces.

HIV epidemics result in selection bias

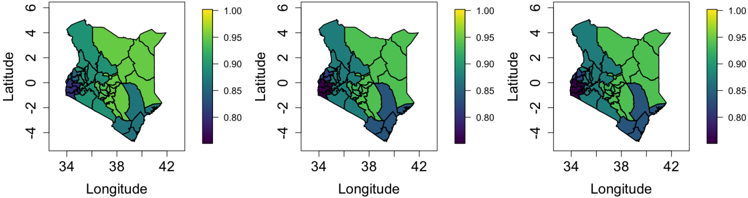


Figure 6: Maps of HIV adjustment ratios of reported U5MRs to "true" U5MRs, that is (2), by survey, in 1995. The 3 columns represent the adjustments from the 2003, 2008–2009, 2014 surveys.

A Smoothed Direct Model

Following Mercer et al. (2015) we use a hybrid model for small-area estimation (SAE): we will refer to this as the smoothed direct model.

Again, the key step is to take as likelihood the asymptotic sampling distribution of a suitable estimator.

Let

- y_{it} be the logit of the U5MR weighted estimator ${}_{60}\widehat{q}_{0,it}$ and
- $\widehat{V}_{\text{DES},it}$ the design-based variance

in area i, period t.

The Smoothed Direct Model

Hierarchical Model:

1. The Data Model:

$$\underbrace{y_{it} \mid \lambda_{it} \sim \mathrm{N}\left(\lambda_{it}, \widehat{V}_{\text{DES},it}\right)}_{\text{Survey design acknowledged here}} \qquad .$$

2. The Space-Time (Random Effects) Prior:

$$\underbrace{\lambda_{it} = f(\text{ space } i, \text{ time } t)}_{\text{Smoothing here}}.$$

A Smoothed Direct Model

Fitting (so-far) carried out in `R` using the `survey` and `INLA` packages, these are wrapped in the `SUMMER` package, along with other plotting and data preparation functions.

Current implementation:

- Modeled in discrete time (with random walk (RW) models),
- Modeled over discrete space (with ICAR models),
- Independent space-time interaction terms.

Aim is for a simple, transparent, robust model.

Model

The data model is

$$y_{it} | \lambda_{it} \sim \text{N}(\lambda_{it}, \widehat{V}_{\text{DES},it}),$$

where

- y_{it} is the logit of the direct estimator in area i and period t,
- λ_{it} is the logit of the true U5MR in county i and period t, and we emphasize that $\widehat{V}_{\text{DES},it}$ is known.
- Important point: Any estimate can be added to the totality of data in this way, so long as it has an associated standard error.

We decompose λ_{it} into temporal, spatial and space-time components:

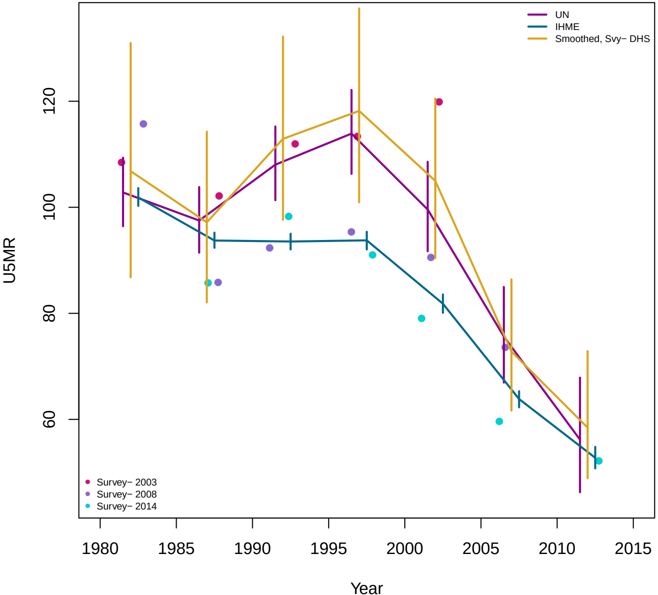
$$\lambda_{it} = \underbrace{\mu}_{\text{Intercept}}$$

$$+ \underbrace{\alpha_t}_{\text{Independent}} + \underbrace{\gamma_t}_{\text{Random Walk}} \qquad \text{Temporal Model}$$

$$+ \underbrace{\theta_i}_{\text{Independent}} + \underbrace{\phi_i}_{\text{ICAR}} \qquad \text{Spatial Model}$$

$$+ \underbrace{\delta_{it}}_{\text{Interaction}} \qquad \text{Space-Time Model}$$

Sanity check of model fit at the national level over time



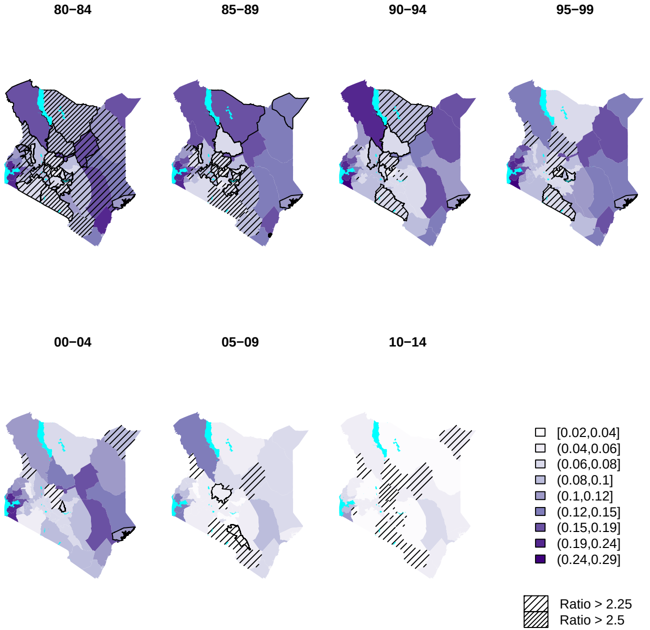


Figure 8: Smoothed estimates at the Admin 1 level.

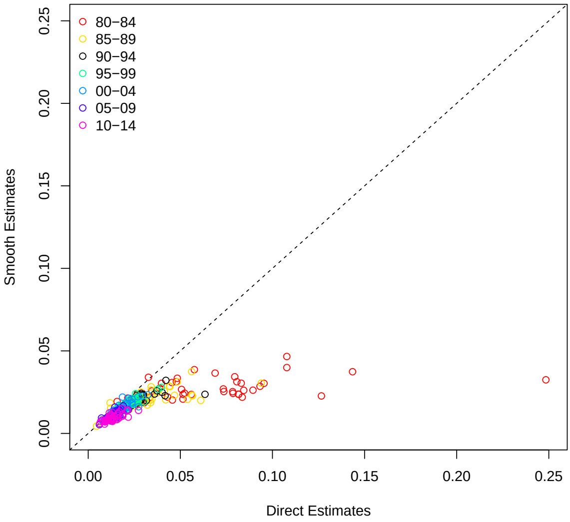
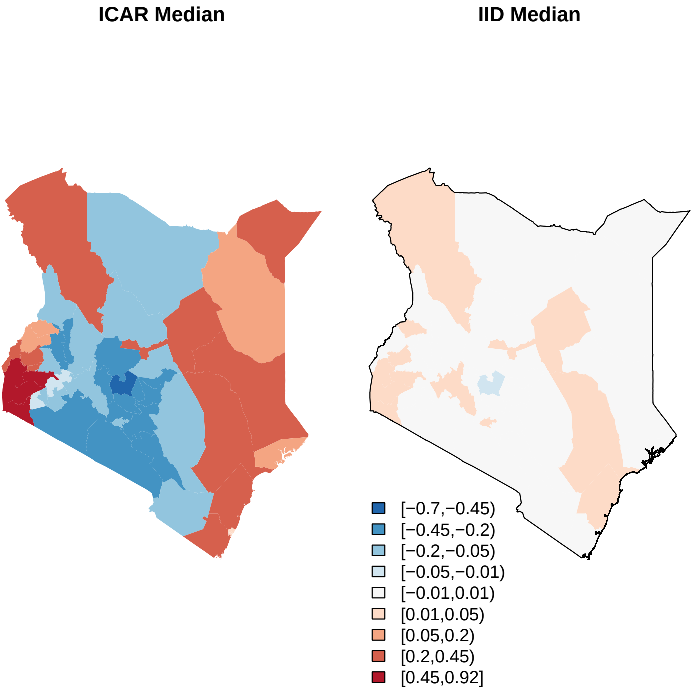


Figure 9: Posterior standard deviations of U5MR estimates versus standard errors of direct (weighted) estimates.


ICAR Median
IID Median
[−0.7,−0.45)
[−0.45,−0.2)
[−0.2,−0.05)
[−0.05,−0.01)
[−0.01,0.01)
[0.01,0.05)
[0.05,0.2)
[0.2,0.45)
[0.45,0.92]

How Does the Variation Apportion?

	Median	Proportion
RW2 (Time)	0.132	40.1
ICAR (Space)	0.130	39.5
Space Unstructured	0.004	1.2
Time Unstructured	0.004	1.1
Time by Space Interaction	0.059	18.0

Table 1: Proportion of variation contributed by random effects at Admin 1 level.

Very large temporal and spatial contributions to the variation.

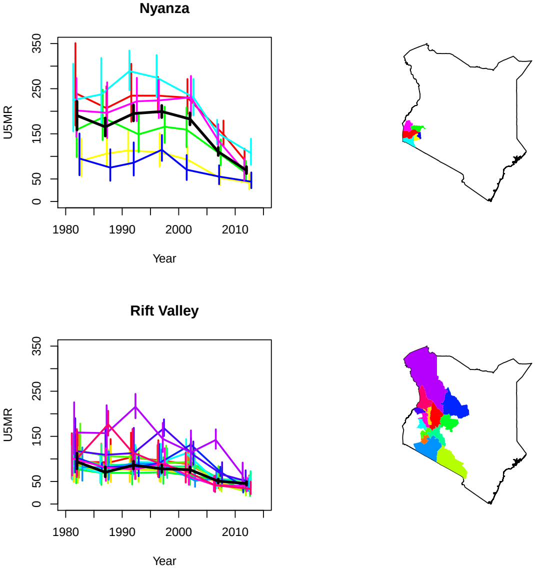


Figure 10: Smoothed regional U5MR estimates for Nyanza and Rift Valley from space-time-smoothing model with estimates from constituent Admin 1 county areas.

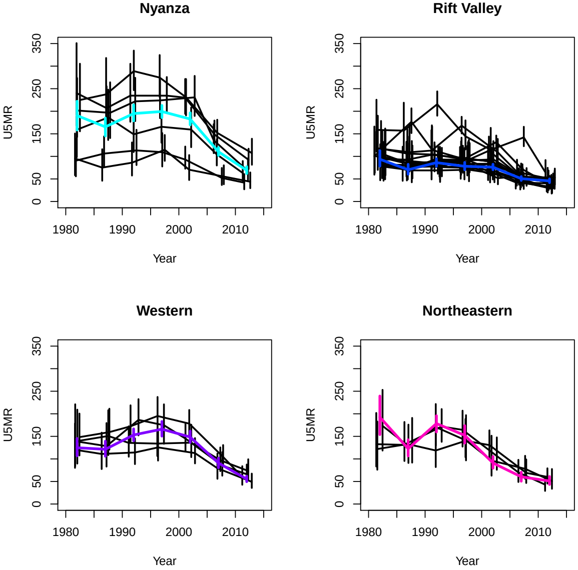


Figure 11: Admin 1 estimates within regions.

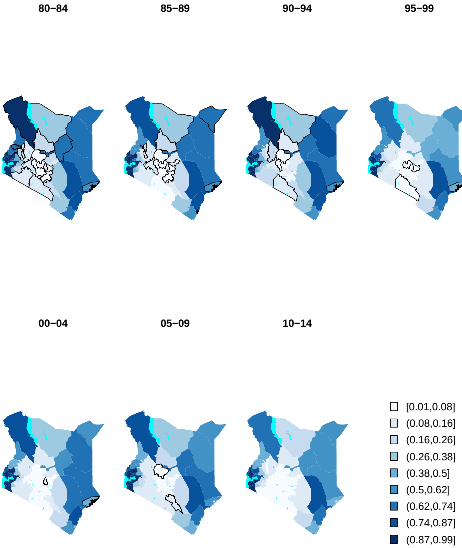


Figure 12: Posterior probability of U5MR exceeding 10%.

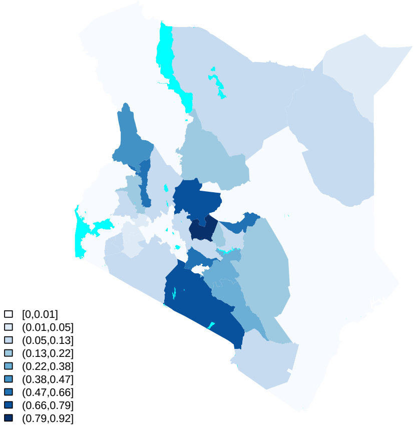


Figure 13: Posterior probability that U5MR is less than 33 deaths per 1,000 births (MDG4 target).

Scaling Up (Li et al., 2018)

We have used this model for 35 African countries, with Type IV (Knorr-Held, 2000) interactions (RW2 $\times$ ICAR).

Spatial scale is Admin 1 and temporal scale is 5-Year periods for data, 1-year periods for estimates.

Data:

- 121 DHS in 35 countries
- 1.2 million children
- 192 million child-months

UN have endorsed these estimates.

Takes around 2.5 hours to obtain estimates for all countries – separate models for each country.

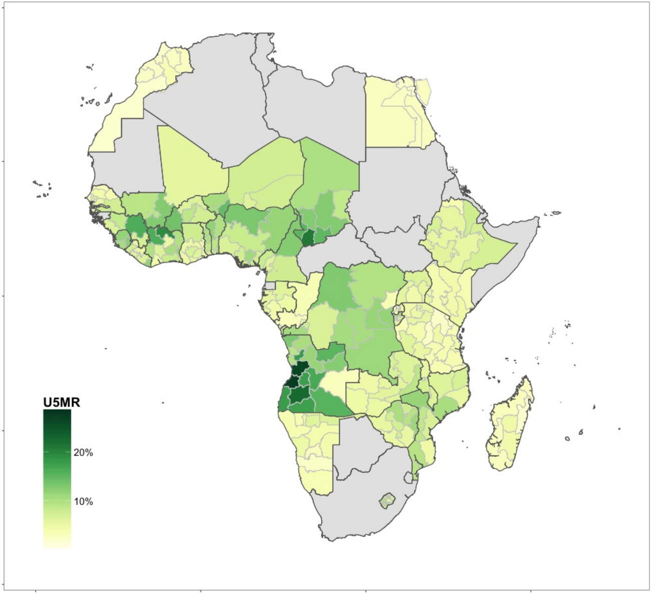


Figure 14: Predictions of U5MR for 2015, in 35 countries of Africa.

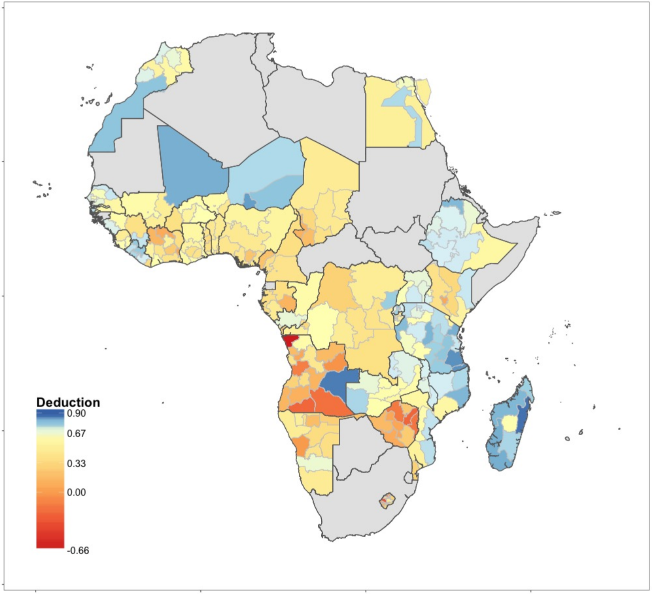


Figure 15: Percentage reduction from 1990 to 2015, in 35 countries of Africa.

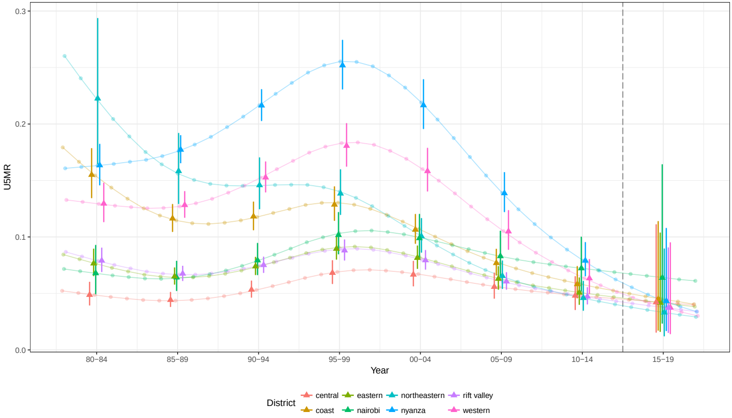


Figure 16: Posterior median estimates for Kenya districts.

Estimation at the Pixel Level

Pixel Surfaces

It is now common to construct spatial surfaces of demographic and health indicators at the "pixel" level:

- Population (Wardrop et al., 2018).
- Malaria (Gething et al., 2016).
- U5MR (Golding et al., 2017).
- Vaccination (Utazi et al., 2018)
- HIV testing in women; stunting in children; anemia in children; household access to improved sanitation (Gething et al., 2015).
- Child growth failure (Osgood-Zimmerman et al., 2018).
- Educational attainment (Graetz et al., 2018).
- …

These maps are based, in large part, on data from surveys, i.e, DHS, MICS,...

Small Area Estimation

In traditional SAE the aim is to estimate true counts or population averages (e.g., fraction with disease) over a group of domains (areas).

Data arise from surveys, often with a complex design.

Areas historically correspond to administrative regions (in which people live) rather than pixel regions (in many of which, nobody lives).

Traditional SAE (Rao and Molina, 2015) does not emphasize spatial smoothing, so no accepted approach as yet (at least not amongst the statistical community...).

The groups who are producing pixel-level maps, almost universally use geostatistical models, which are often referred to as Gaussian process (GP) models.

A GP Spatial Model

Suppose we have n cluster locations $\boldsymbol{s}_i$, $i = 1, \ldots, n$, at which data is collected.

Basically, GP models assume that

$$\boldsymbol{S} = (S_1, \ldots, S_n)$$

arise from a zero mean multivariate normal distribution with variances

$$\text{var}(S_i) = \sigma_s^2$$

and correlations $\text{corr}(S_i, S_j)$.

The obvious approach in a spatial setting is to assume a form such that the correlation between S_i and S_j decreases as d_{ij}, the distance between the locations at which S_i and S_j are measured, decreases.

A model in which the correlations are a function of distance only between the points is known as isotropic.

A GP Spatial Model

In its simplest form, the GP model has two parameters, σ_s^2, which determines the scale of the spatial variability, and ρ, which determines the extent of the spatial variability.

A simple form is,

$$\text{corr}(S_i, S_j) = \rho^{d_{ij}}$$

where

- $d_{ij} = ||\boldsymbol{s}_i - \boldsymbol{s}_j||$ is the distance between the centroids of areas i and j, and
- $\rho > 0$ is a parameter that determines the extent of the correlation; ρ is the correlation between the residual spatial variability in two locations that are one unit of distance apart.

The correlation above is the marginal correlation between the random variables S_i and S_j.

A GP Spatial Model

More generally, the correlations can be modeled as a Matérn correlation function (Stein, 1999):

$$\text{corr}(S_i, S_j) = \frac{1}{\Gamma(\upsilon + 1/2)(4\pi)^{1/2}\kappa^{2\upsilon}2^{\upsilon-1}}(\kappa d_{ij})^{\upsilon} K_{\upsilon}(\kappa d_{ij})$$

where $K_{\upsilon}(\cdot)$ is a modified Bessel function of the second kind, $\kappa > 0$ is a scale parameter and $\upsilon > 0$ is a smoothness parameter.

In general, difficult to estimate many parameters in a spatial model and often υ is fixed.

Requires estimation of a spatial variance parameter and an effective range.

A GP Spatial Model

The multivariate model with correlations of this form is computationally expensive to fit, because one has to carry out operations on the $n \times n$ covariance matrix, which we call $\boldsymbol{\Sigma}$.

The multivariate normal distribution $\boldsymbol{S}|\boldsymbol{\Sigma} \sim \mathrm{N}(\mathbf{0}, \boldsymbol{\Sigma})$ is given by

$$p(\boldsymbol{S}) = (2\pi|\boldsymbol{\Sigma}|)^{-1/2} \exp\left(-\frac{1}{2}\boldsymbol{S}^{\mathsf{T}}\Sigma^{-1}\boldsymbol{S}\right),$$

so to evaluate the density we need to calculate a determinant and an inverse.

The covariance matrix $\boldsymbol{\Sigma}$ depends on the parameters of the spatial covariance function.

We now show how this model is used in the context of U5MR estimation.

Model-Based Geostatistics with a GP Prior

For simplicity consider a binary outcome and let Y_{ik} be the number of individuals out of n_{ik} with the characteristic of interest in cluster k of area i.

Wakefield et al. (2018) describe the geostatistics model:

$$\begin{aligned} Y_{ik} | \theta_{ik} &\sim \text{Binomial}(n_{ik}, \theta_{ik}) \\ \log\left(\frac{\theta_{ik}}{1-\theta_{ik}}\right) &= \beta_0 + \gamma I(\boldsymbol{s}_{ik} \in \text{ urban }) + \beta x_{ik} + \epsilon_{ik} + S_{ik}^{\text{CONT}} \end{aligned}$$

where

- $\theta_{ik} = \theta(\boldsymbol{s}_{ik})$ is the risk at location $\boldsymbol{s}_{ik}$,
- γ describes the association with urban,
- x_{ik} are covariates,
- $\epsilon_{ik} \sim \text{N}(0, \sigma_\epsilon^2)$ is the nugget,
- S_{ik}^{CONT} are spatial random effects, assumed to arise from a Gaussian process.

Model-Based Inference

Alternatively a discrete spatial model can be used:

$$\log\left(\frac{\theta_{ik}}{1-\theta_{ik}}\right) \quad = \quad \beta_0 + \gamma I(\boldsymbol{s}_{ik} \in \text{ urban }) + \beta x_{ik} + \epsilon_{ik} + S_i^{\text{DISC}}$$

where

- S_i^{DISC} are discrete spatial random effects that follow an ICAR (Markov Random Field) model (Besag et al., 1991).

Model-Based Inference

For either model, area estimates are obtained by averaging point estimates with respect to the population from:

$$\theta_i = \frac{\int_s \theta(\boldsymbol{s}) d(\boldsymbol{s})\, d\boldsymbol{s}}{\int_s d(\boldsymbol{s})\, d\boldsymbol{s}}$$

where $d(\boldsymbol{s})$ is population density at location $\boldsymbol{s}$.

In practice, the continuous spatial model is always approximated by some form of discretization, so the integral is approximated by summing over a grid.

We need to know all the covariates and urban/rural status of everywhere on the grid.

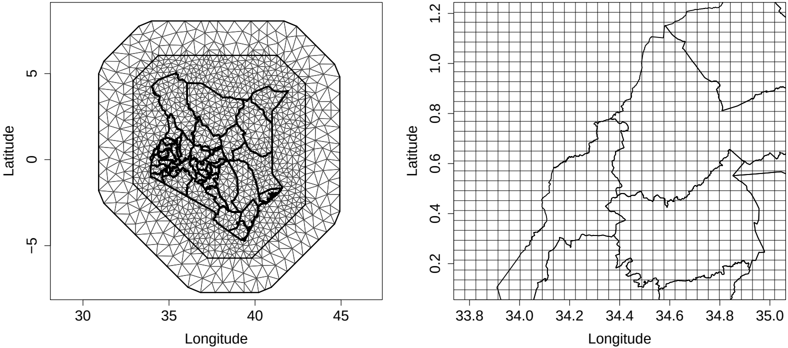


Figure 17: Mesh on which SPDE calculations are carried out (top left), zoomed in grid on which predictions are performed (right).

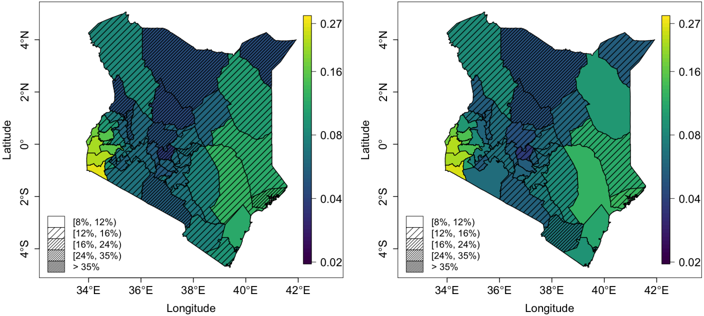


Figure 18: Kenya U5MR estimates in 2000 using discrete spatial model (left), and continuous spatial model (right).

Point estimates are very similar, but more uncertainty associated with the discrete spatial model estimates.

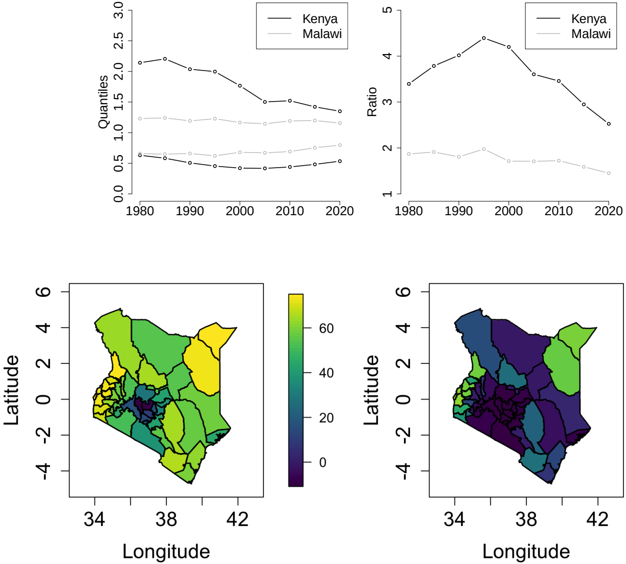


Figure 19: Top row: Kenya and Malawi within-country variability in U5MR (5% and 95% quantiles of pixel distribution). Bottom row: percentage drop from 1990–2015 (left), posterior probability of attaining MDG goal (right).

Comparison of Discrete and Continuous Spatial Models

MSE comparison based on 400 (out of 1600) clusters from 2014 Kenya DHS.

Let:

- $Y_{ip}^{(1)}$ denote the weighted estimator.
- $Y_{ip}^{(2)}$ the smoothed estimator from continuous space model.
- $Y_{ip}^{(3)}$ the smoothed estimator from discrete space model: ICAR $\times$ AR(1), with the latter having yearly resolution,

$p = \{$1990–1994,1995–1999, 2000–2004, 2005–2009, 2010–2014 $\}$, $j = 1, 2, 3$.

We compare these estimates with the weighted estimates from (approximately) 1200 (left-out) clusters from 2014, y_{ip} (the "truth").

In particular, we calculate,

$$\text{MSE}_p^{(j)} = \frac{1}{47} \sum_{i=1}^{47} \left(Y_{ip}^{(j)} - y_{ip} \right)^2 . \quad (3)$$

MSE Comparison

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.

Conclusions:

- Spatial models have very similar predictive ability, with the continuous model being slightly more accurate.
- Both show a dramatic improvement over the weighted estimates.

Acknowledging the Complex Survey Design

Statistical Issues with Complex Sampling

Ignoring the design leads to the possibility of:

- Bias (if stratification variables are associated with the outcome).
- An inappropriate measure of variance (cluster sampling breaks independence of outcomes).

We report on a limited simulation exercise that investigates the impact of ignoring the design.

As a simple example, suppose the strata are urban/rural.

If we ignore this aspect then

- area-level estimates will be biased unless:
  - the outcome does not depend on strata membership, or
  - sampling of strata is in the same proportion as the population frequencies (so not stratified!).
- pixel-level estimates will be biased unless:
  - the outcome does not depend on strata membership.

Note: If population density and/or travel time are in the covariate model, may get partial correction.

Model-Based Geostatistics

It has become the norm to ignore stratification and assume the geostatistics model:

$$\begin{aligned} Y_{ik}|\theta_{ik} &\sim \text{Binomial}(n_{ik}, \theta_{ik}) \\ \log\left(\frac{\theta_{ik}}{1-\theta_{ik}}\right) &= \beta_0 + \beta x_{ik} + \epsilon_{ik} + S_{ik}^{\text{CONT}}. \end{aligned}$$

All of the pixel created map references given earlier ignore urban/rural...

Gething and Burgert-Brucker (2017) reported mixed accuracy for different outcomes using this model (poor for vaccination surfaces, for example).

Accounting for Complex Sampling

We consider the simplified situation in which we have:

- A single survey.
- A binary outcome.

Using Kenya geography, we simulate a single complete population:

- Clusters: 96,251 enumeration areas (EAs), 32% are urban.
- Strata used in DHS in 2014 are 47 counties and urban/rural (92 in total, Nairobi and Mombasa are entirely urban).
- From the Kenya 2014 DHS report we know the numbers of urban/rural EAs by district and we match these numbers by thresholding on a population density surface.
- Within each EA, assume 25 households, with one mother in each household and one birth per mother.

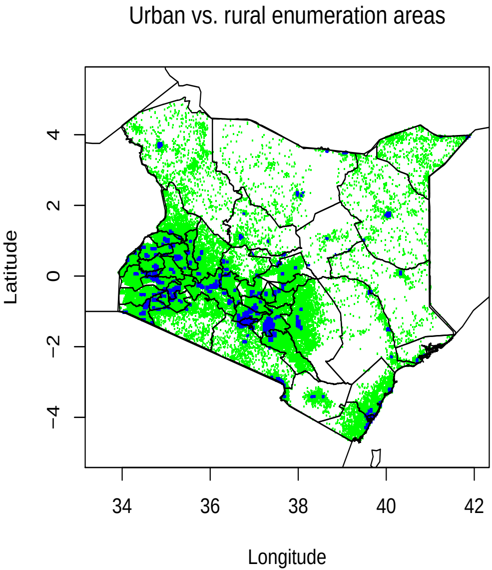


Figure 20: Sampling frame for Kenya simulation.

Accounting for Complex Sampling

We have $n_j = 25$ births at each EA (cluster) location $\boldsymbol{s}_j$, $j = 1, \dots, n$, and we generate neonatal deaths Y_j according to

$$\begin{aligned} Y_j | \theta(\boldsymbol{s}_j) &\sim \text{Binomial}\left(n_j, \theta(\boldsymbol{s}_j)\right) \\ \log\left(\frac{\theta(\boldsymbol{s}_j)}{1-\theta(\boldsymbol{s}_j)}\right) &= \beta_0 + \gamma I(\boldsymbol{s}_j \in \text{ urban }) + \epsilon_j + S(\boldsymbol{s}_j), \end{aligned}$$

where

- $\epsilon_j \sim_{iid} \text{N}(0, \tau^2)$ (the nugget),
- $S(\boldsymbol{s})$ is a Gaussian Process (GP) with Matérn covariance function and (effective) range ϕ and variance σ^2.

The nugget term induces within-cluster dependence.

Accounting for Complex Sampling

Assume inference is at the county level.

Methods to be compared:

- Naive: Assume binomial (unweighted) counts in each county. This gives an estimate $\widehat{\theta}_i^{\text{BIN}}$ and a variance from which an asymptotic CI can be calculated.
- Direct estimates: This gives an estimate $\widehat{\theta}_i^{\text{DIR}}$ and a variance from which an asymptotic CI can be calculated.

Accounting for Complex Sampling

- Smoothed Direct: Take logit of direct estimates θ_i^{DIR} with appropriate design-based estimator and model as Mercer et al. (2015),

$$\begin{aligned} \text{logit}(\widehat{\theta}_i^{\text{DIR}}) &\sim \text{N}(\eta_i, \widehat{V}_i) \\ \eta_i &= \beta_0 + \underbrace{\epsilon_i}_{\text{Independent}} + \underbrace{S_i}_{\text{ICAR}} \end{aligned}$$

County smoothed direct estimate

$$\widehat{\theta}_i^{\text{SDIR}} = \text{expit}(\widehat{\beta}_0 + \widehat{\epsilon}_i + \widehat{S}_i).$$

Accounting for Complex Sampling

- Smoothed Adjusted Discrete Spatial Model at the cluster level:

$$\begin{aligned} Y_j|\theta_j &\sim \text{Binomial}(n_j, \theta_j) \\ \text{logit}(\theta_j) &= \beta_0 + \gamma I(\boldsymbol{s}_j \in \text{ urban }) + \underbrace{\epsilon_{i[j]}}_{\text{Independent}} + \underbrace{S_i}_{\text{ICAR}} + \underbrace{\delta_j}_{\text{Independent}}. \end{aligned}$$

  Obtain 2 estimates for each county i:

$$\begin{aligned} \widehat{\theta}_{i1} &= \text{expit}(\widehat{\beta}_0 + \widehat{\epsilon}_i + \widehat{S}_i) \\ \widehat{\theta}_{i2} &= \text{expit}(\widehat{\beta}_0 + \widehat{\gamma} + \widehat{\epsilon}_i + \widehat{S}_i) \end{aligned}$$

  Then

$$\widehat{\theta}_i = q_i \widehat{\theta}_{i1} + (1 - q_i)\widehat{\theta}_{i2}$$

  where q_i is the proportion of the births that occur in rural clusters.
- Smoothed Adjusted Continuous Spatial Model at the cluster level:

$$\begin{aligned} Y_j|\theta_j &\sim \text{Binomial}(n_j, \theta_j) \\ \text{logit}(\theta_j) &= \beta_0 + \gamma I(\boldsymbol{s}_j \in \text{ urban }) + \underbrace{\epsilon_j}_{\text{Independent}} + \underbrace{S_j}_{\text{GP}} \end{aligned}$$

Accounting for Complex Sampling

Methods comparison: bias, MSE, Average of Variance, 80% CI coverage.

Parameters (in all simulations):

- $\beta_0 = -2$, $\gamma = -0.5$ (so urban lower)
- $\sigma^2 = 0.15^2$, effective range $\phi = 300\,\text{km}$, $\tau^2 = 0.1^2$.

Two simulations:

1. Unstratified sampling.
2. Stratified sampling in which we oversample urban clusters. Specifically, in each county sample twice as many urban as rural clusters.

Preliminary Results[1]

- Unstratified sampling:

Method	Bias	MSE	Ave. Var.	80% coverage
Naive	-0.020	0.060	0.051	0.78
Direct estimates	-0.020	0.060	0.053	0.75
Smoothed Direct	0.012	0.018	0.018	0.78
Discrete Spatial	-0.014	0.011	0.015	0.84
Continuous Spatial	-0.005	0.012	0.010	0.72

- Stratified sampling:

Method	Bias	MSE	Ave. Var.	80% coverage
Naive	-0.082	0.069	0.053	0.75
Direct estimates	-0.029	0.066	0.058	0.73
Smoothed Direct	0.005	0.021	0.020	0.78
Discrete Spatial	-0.015	0.011	0.016	0.86
Continuous Spatial	-0.005	0.012	0.010	0.72

[1] Bias is logit $\widehat{\theta}_i$ − logit θ_i where θ_i is truth

Model Validation

Model Validation

No consensus on how to validate model, cross-validation is the most common approach, but details on how splits were made often sketchy, as are exact ways in which predictions obtained (supplementary materials hide many sins...).

When bias is reported, what is the "truth"?

By construction, spatial models smooth the covariate mean in areas with no data.

Wakefield et al. (2018) compared predictions for U5MR in Kenya from discrete and continuous spatial models:

- "Truth" (direct estimates with small variance) is only available at Admin-1, 5-year scale.
- Discrete and continuous models performed equally well, but below Admin-1, who knows?

Now investigating the use of proper scoring rules (Gneiting and Raftery, 2007).

Covariate Modeling

Distinguish between:

- Individual-level modeling, for example, for U5MR, Balk et al. (2004).
- Surface modeling, in which we require covariates to be available at all prediction points.

Some approaches:

- Often some kind of backward elimination (e.g., Utazi et al., 2018) or all subsets (e.g., Gething et al., 2015).
- Stacked generalization/super learner (Bhatt et al., 2017; Golding et al., 2017).

In general, inference/uncertainty estimates do not correctly account for the selection of the final covariate model.

Discussion

Discussion: Comparison of Models

	Direct Estimation	Smoothed Direct	Discrete Spatial	Continuous Spatial
Robustness	✓✓✓✓	✓✓✓	✓✓	✓
Transparency	✓✓✓✓	✓✓✓	✓✓	✓
Sparse Data	✓	✓✓	✓✓✓✓	✓✓✓✓
Spatial Scale	✓	✓	✓✓✓✓	✓✓✓✓
Data Required	✓✓✓✓	✓✓✓✓	✓✓✓	✓✓
Flexibility	✓	✓✓	✓✓✓	✓✓✓✓

Table 3: Comparison of approaches to SAE.

General strategy: See if estimates from different models are consistent with each other.

There is some skepticism of even national estimates (e.g., Boerma et al., 2018), let alone SAE or pixel level estimation.

Discussion

Substantive:

- Follow-up to Admin-1 in sub-Saharan Africa paper: Admin-2 including summary birth history data.
- Asia at Admin-1.
- Examination of biases in DHS data.
- Measles: modeling vaccination coverage and spatio-temporal disease count data.

Methodological:

- Consensus on estimation at the pixel level.
- Modeling summary birth history.
- Examination of implications of ignoring the design.
- Points/polygons problem (Wilson and Wakefield, 2018).
- Examination of model validation techniques.
- Covariate modeling (how to use information on conflicts?).
- Spatial APC models with survey data.

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.

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.

Binder, D. (1983). On the variances of asymptotically normal estimators from complex surveys. *International Statistical Review 51*, 279–292.

Boerma, T., C. Victora, and C. Abouzahr (2018). Monitoring country progress and achievements by making global predictions: is the tail wagging the dog? *The Lancet*. Published Online, April 13, 2018.

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.

Gething, P. W. and C. R. Burgert-Brucker (2017). The DHS program modeled map surfaces: understanding the utility of spatial interpolation for generating indicators at subnational administrative levels.

Gething, P. W., D. C. Casey, D. J. Weiss, D. Bisanzio, S. Bhatt, E. Cameron, K. E. Battle, U. Dalrymple, J. Rozier, P. C. Rao, et al. (2016). Mapping plasmodium falciparum mortality in africa between 1990 and 2015. *New England Journal of Medicine 375*, 2435–2445.

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

References III

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 390*, 2171–2182.

Graetz, N., J. Friedman, A. Osgood-Zimmerman, R. Burstein, M. H. Biehl, C. Shields, J. F. Mosser, D. C. Casey, A. Deshpande, L. Earl, et al. (2018). Mapping local variation in educational attainment across Africa. *Nature 555*, 48.

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

Li, R., Y. Hsiao, J. Godwin, B. B. Martin, J. Wakefield, and S. Clark (2018). 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. *Submitted*.

References IV

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.

Osgood-Zimmerman, A., A. I. Millear, R. W. Stubbs, C. Shields, B. V. Pickering, L. Earl, N. Graetz, D. K. Kinyoki, S. E. Ray, S. Bhatt, et al. (2018). Mapping child growth failure in africa between 2000 and 2015. *Nature 555*, 41.

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

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

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

References V

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

Wakefield, J., G.-A. Fuglstad, A. Riebler, J. Godwin, K. Wilson, and S. J. Clark (2018). Estimating under five mortality in space and time in a developing world context. *Statistical Methods in Medical Research*. To Appear.

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.

Wardrop, N., W. Jochem, T. Bird, H. Chamberlain, D. Clarke, D. Kerr, L. Bengtsson, S. Juran, V. Seaman, and A. Tatem (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.

Wilson, K. and J. Wakefield (2018). Pointless continuous spatial surface reconstruction. *Biostatistics*. To appear.

Technical Appendix: Individual versus Ecological Modeling

Individual versus Ecological Modeling

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 individual variables.

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.

Individual versus Ecological Modeling

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.

If we wish to obtain predictions for unobserved locations on the basis of a covariate model, then those covariates must be available.

Technical Appendix: Random Walk Models

RW1 model for temporal dependence

We describe a particular limiting autoregressive model that is a popular tool for nonparametric smoothing.

The model takes the limit of the AR1 model as $\rho \to 1$ and takes the form

- *Stage 1:* $Y_t = \mu_t + T_t + \epsilon_t$, $\epsilon_t \sim_{iid} \mathrm{N}(0, \sigma_\epsilon^2)$.
- *Stage 2:* $T_t = T_{t-1} + \tau_t$, $\tau_t \sim_{iid} \mathrm{N}(0, \sigma_\tau^2)$.

This is known as a random walk model of order one, which we write as RW1.

Note: depends on a single parameter σ_τ^2.

RW1 Model

The undirectional version is

$$T_t | T_{t-1}, T_{t+1} \sim \mathrm{N}\left(\frac{1}{2}(T_{t-1} + T_{t+1}), \frac{\sigma_\tau^2}{2}\right),$$

for $1 < t < n$.

For prediction, future values have the conditional distribution:

$$T_{n+s} | T_1, \ldots, T_n, \sigma_\tau^2 \sim \mathrm{N}\left(\underbrace{T_n}_{\text{Predictive Mean}}, \underbrace{s \times \sigma_\tau^2}_{\text{Predictive Variance}}\right),$$

for $s > 0$.

Hence, predictions into the future have the same level, and the variance is linear in s.

RW2 Model

For the RW1 model, a least squares fit to the two adjacent points T_{t-1} and T_{t+1} gives a fitted mean of $\frac{1}{2}(T_{t-1} + T_{t+1})$.

The RW2 model gives more smoothing by smoothing over 4 neighbors.

The undirectional version is

$$T_t | T_{t-1}, T_{t-2} T_{t+1}, T_{t+2} \sim \mathrm{N}\left\{ \frac{4}{6}(T_{t+1} + T_{t-1}) - \frac{1}{6}(T_{t+2} + T_{t-2}), \frac{\sigma_\tau^2}{6} \right\}$$

for $1 < t < n$.

A least squares fit of a quadratic model to the four adjacent points $T_{t-2}, T_{t-1}, T_{t+1}, T_{t+2}$ gives the above fitted mean,
i.e. $\frac{4}{6}(T_{t+1} + T_{t-1}) - \frac{1}{6}(T_{t+2} + T_{t-2})$.

RW2 Model

For prediction, future values have the conditional distribution:

$$T_{n+s}|T_1,\dots,T_n,\sigma_\tau^2 \sim \mathrm{N}\left\{\underbrace{(1+s)T_n - sT_{n-1}}_{\text{Predictive Mean}}, \underbrace{(1+2^2+\cdots+s^2)\times\sigma_\tau^2}_{\text{Predictive Variance}}\right\},$$

for $s > 0$.

Hence, the temporal trend is determined by the last two points, and we have a linear trend.

So the trend is more flexible, but the variance is larger, when compared to the RW1 model.

RW1 and RW2 Models

Figure **??** shows simulated data from a sine curve and then fit using RW1 and RW2 models.

The resultant fits are in indicated.

Note that the RW2 fit is smoother.

The RW1 and RW2 models are usually fitted using a Bayesian approach; the prior on σ_τ^2 can be used to control the amount of smoothing:

- Giving greater weight to smaller (larger) values of σ_τ^2 gives more (less) smoothing.
- In the limit as $\sigma_\tau^2 \to 0$, the RW1 model tends to a horizontal line, and the RW2 model tends to a linear trend in time.

RW1 and RW2 Models

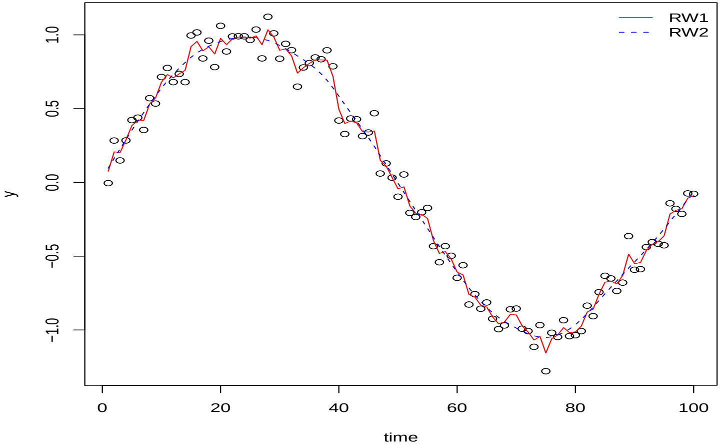


Figure 21: Simulated data and RW1 and RW2 fits.