Bayesian Subnational Estimation using Complex Survey Data: Spatial Models for Survey Data

Jon Wakefield

Departments of Statistics and Biostatistics
University of Washington

Outline

Motivation

Motivation

- We are interested in examining subnational variation in health and demographic indicators across some study region.
- This can be achieved by prevalence mapping, which can produce continuous prevalence surfaces, or area-level maps.
- We take as aim, obtaining area-level estimates in order to target resources and examine progress towards goals, which is explicitly the objective of small area estimation (SAE).
- In the terminology of survey sampling, SAE is an example of domain (sub-population) estimation.
- "Small" refers to the fact that we will typically base our inference on a small sample from each area, so it is not a description of geographical size, in the limit there may some areas in which there are no data.

Small Area Estimation

- Consider a study region partitioned into n disjoint and exhaustive areas, indexed by i, $i = 1, \ldots, n$.
- As a concrete example, suppose we are interested in a particular condition so that the response is a binary outcome, Y_{ik}, for $k = 1, \ldots, N_i$, individuals in area i.
- Based on samples that are collected in the areas (though some areas may contain no samples), common targets of estimation are:
  - The population totals:

$$T_i = \sum_{k=1}^{N_i} Y_{ik}.$$

  - The prevalence of the condition in each area:

$$P_i = \frac{1}{N_i} \sum_{k=1}^{N_i} Y_{ik} = \frac{T_i}{N_i}.$$

Background Reading on SAE

- The classic text on SAE is Rao and Molina (2015); not the easiest book to read, and little material on spatial smoothing models.
- An excellent review of SAE is Pfeffermann (2013), though again little on spatial modeling.
- The SAE literature distinguishes between direct estimation, in which data from the area only is used to provide the estimate in an area, and indirect estimation, in which data from other areas are used to provide the estimate.

Overview of Modeling

- We first briefly review more traditional SAE techniques.
- We then describe two approaches to spatial and space-time modeling of survey data:
  - A smoothed direct method, that treats the direct (weighted) estimates as data, and then fits a spatial smoothing model.
  - A cluster-specific model with a (say) overdispersed sampling model at the cluster level, and a smoothing model being applied to the prevalence – both discrete and continuous spatial models will be discussed.
- We also describe the discrete hazards model that is used for producing subnational estimates of of the under-5 mortality risk (U5MR).

Traditional Methods for SAE

Design-Based Inference Based on Weighted Estimators

- Suppose we undertake a complex design and obtain outcomes y_{ik} in area i, $k \in S_i$, where S_i is the set of samples that were in area i.
- Along with the outcome, there is an associated design weight w_{ik}.
- Under the design-based approach to inference, it is common to use the weighted estimator of the prevalence:

$$\widehat{p}_i = \frac{\sum_{k \in S_i} w_{ik} y_{ik}}{\sum_{k \in S_i} w_{ik}}.$$

- A variance, $\widehat{V}_i$, appropriate for the design, may be calculated, either analytically, or through resampling techniques such as the jackknife.
- Asymptotically (that is, in large samples):

$$\widehat{p}_i \sim \mathrm{N}(p_i, \widehat{V}_i).$$

Direct Estimation

- The simplest approach is to simply map the direct estimates $\widehat{p}_i$.
- To assess the uncertainty, one may map the width of (say) the 95% confidence interval:

$$\widehat{p}_i \pm 1.96 \times \sqrt{\widehat{V}_i}.$$

- If the samples in each area are large, so that $\widehat{V}_i$ is of acceptable size, then this approach works well.
- We would like to carry out some form of smoothing, but in the case of complex survey sampling, how should we proceed?

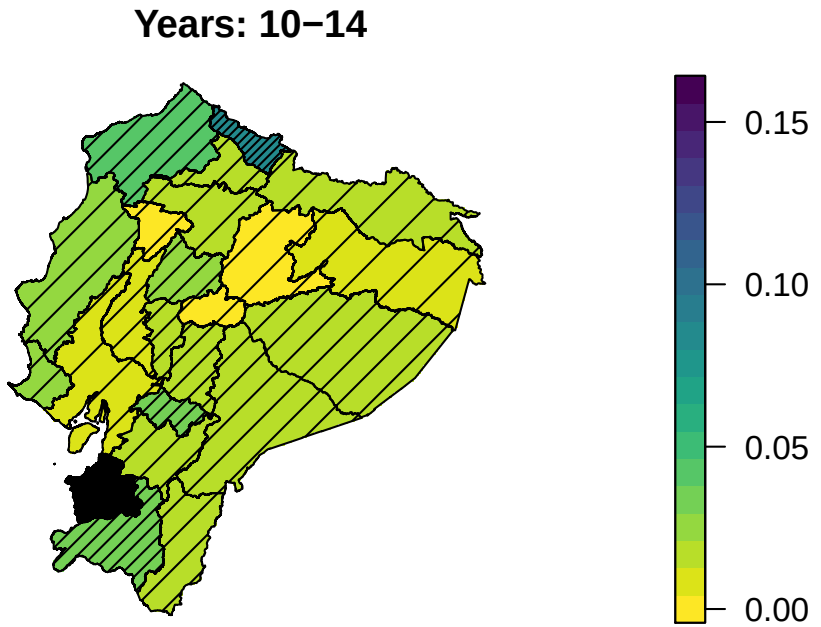


Figure 1: Direct estimates of U5MR in Ecuador, for 2010–2014. Denser hatching indicates greater uncertainty.

- Many approaches have been suggested to obtain estimators with greater precision – we discuss three, to give a flavor.

Synthetic Estimator

- The synthetic estimator is,

$$\widehat{Y}_i^{\text{SYN}} = \frac{1}{N_i} \sum_{k=1}^{N_i} \boldsymbol{x}_{ik}^{\mathsf{T}} \widehat{B},$$

  for $i = 1, \ldots, n$, where

$$\widehat{B} = \left[\sum_{i=1}^{n} \sum_{k \in S_i} w_{ik} \boldsymbol{x}_{ik}^{\mathsf{T}} \boldsymbol{x}_{ik} \right]^{-1} \sum_{i=1}^{n} \sum_{k \in S_i} w_{ik} \boldsymbol{x}_{ik}^{\mathsf{T}} y_{ik}.$$

- Note: Covariates needed for all of population.
- It is assumed that the regression model is appropriate for all areas.
- An example of an indirect estimator, since information is used from all areas.
- In general gives high precision estimates when n is large, since variance is $O(1/n)$, but there is the possibility of large bias.

Survey-Regression Estimates

- In order to deal with the potential large bias, this bias is estimated and then the estimate is adjusted.
- The resultant survey-regression estimator is,

$$\begin{aligned}\widehat{Y}_i^{\text{S-R}} &= \frac{1}{N_i}\sum_{k=1}^{N_i}\boldsymbol{x}_{ik}^{\mathsf{T}}\widehat{B} + \frac{1}{N_i}\sum_{k\in S_i} w_{ik}(y_{ik} - \boldsymbol{x}_{ik}^{\mathsf{T}}\widehat{B}) \\ &= \widehat{Y}_i^{\text{HT}} + (\overline{\boldsymbol{x}}_i - \widehat{\overline{\boldsymbol{x}}}_i^{\text{HT}})^{\mathsf{T}}\widehat{B}\end{aligned}$$

  where $\widehat{Y}_i^{\text{HT}}$ and $\widehat{\overline{\boldsymbol{x}}}_i^{\text{HT}}$ are the Horvitz-Thompson estimates of the population mean in area i, $\overline{Y}_{Ui}$ and $\overline{\boldsymbol{x}}_i$.
- Variance is unfortunately $O(1/n_i)$, so that large variances will result, when there are small samples in areas.
- This survey-regression estimator is reliable in areas with large n_i.

Composite Estimator

- The composite estimator is of the form

$$\widehat{\overline{Y}}_i^{\text{COM}} = \delta_i \widehat{Y}_i^{\text{S-R}} + (1 - \delta_i) \widehat{Y}_i^{\text{SYN}}$$

  with $0 \leq \delta_i \leq 1$ estimated in such a way that for larger n_i we have larger δ_i – this estimator attempts to balance the bias of the synthetic estimator with the instability of the direct estimator.
- Various possibilities exist for estimation of δ_i, in a design-based framework, see Rao and Molina (2015).
- Rather than proceed along this route, we instead consider model-based spatial SAE approaches.

Smoothed Direct Modeling

Smoothed Direct Estimation

- Fay and Herriot (1979) suggested the following hierarchical model, in a landmark paper.
- Let $\widehat{p}_i$ be the weighted estimator of a prevalence p_i, then consider

$$\widehat{\theta}_i = \text{logit}\,(\widehat{p}_i) = \log\left(\frac{\widehat{p}_i}{1-\widehat{p}_i}\right),$$

  which is on the whole of the real line.
- The "data" is taken to be $\widehat{\theta}_i$ and the sampling model is taken as the asymptotic distribution of the direct estimator:

$$\widehat{\theta}_i \sim \text{N}(\theta_i, \widehat{V}_i),$$

  where $\widehat{V}_i$, the variance of the estimator, is known.
- The prior random effects model is

$$\theta_i = \alpha + \boldsymbol{x}_i^{\mathsf{T}}\boldsymbol{\beta} + e_i,$$

  where $\boldsymbol{x}_i$ are area-level covariates and the random effects $e_i \sim_{iid} \text{N}(0, \sigma_e^2)$.
- The model acknowledges the design and also smooths to a global level – it is straightforward to add spatial random effects.

Smoothed Direct Estimation

- The "data" is taken to be $\widehat{\theta}_i$ and the sampling model is taken as

$$\widehat{\theta}_i \sim \mathrm{N}(\theta_i, \widehat{V}_i),$$

  where $\widehat{V}_i$ is the known variance.
- The prior random effects model is

$$\theta_i = \alpha + \boldsymbol{x}_i^{\mathsf{T}}\boldsymbol{\beta} + e_i + S_i,$$

  with
  - $e_i \sim \mathrm{N}(0, \sigma_e^2)$.
  - $[S_1, \ldots, S_n]^{\mathsf{T}} \sim \mathrm{ICAR}(\sigma_s^2)$.
- Known as an area-level SAE model.
- Mercer *et al.* (2015) and Li *et al.* (2019) considered a space-time version of this model.

Smoothed Direct Estimation

- The area-level SAE model has been used by Gutreuter *et al.* (2019) in the context of estimating HIV prevalence and burden in districts of South Africa, using household survey data.
- Among the covariates considered for the prevalence model were:
  - prevalence estimates from antenatal clinics data,
  - population density,
  - percentages of housing units that were “formal dwellings”,
  - dependency ratio (ratio of the numbers of residents aged 15–64 years to those younger than 15 years and older than 64 years),
  - socio-economic quintile,
  - maternal mortality rate.
- A conditional autoregressive (CAR) spatial model (Marhuenda *et al.*, 2013) was used.

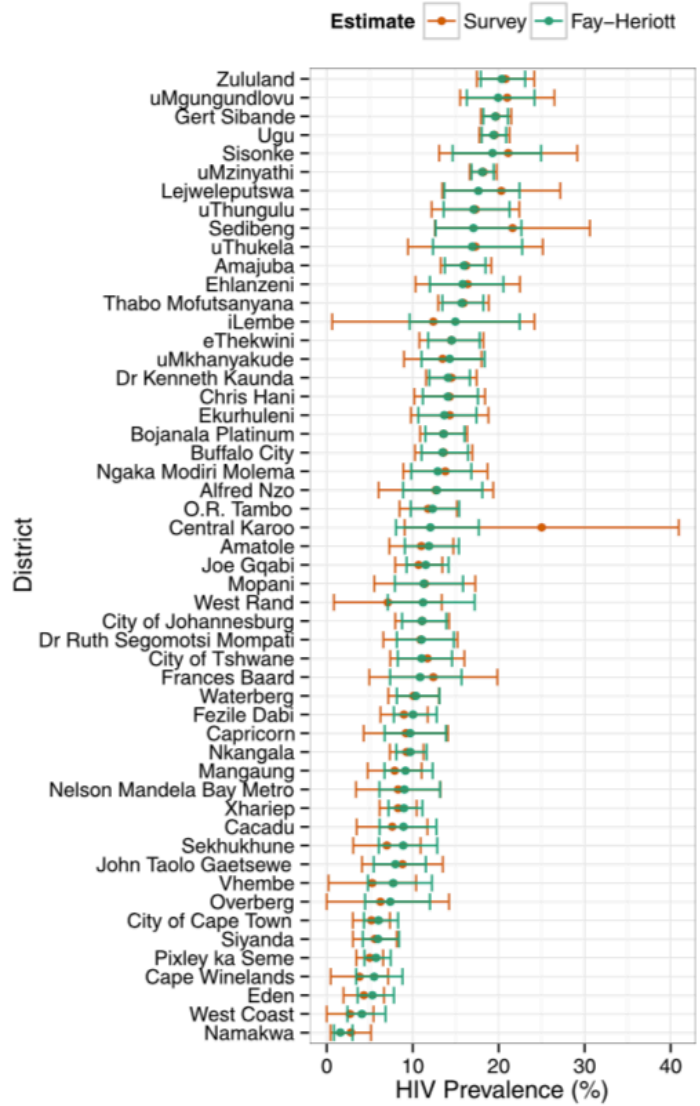


Figure 2: Direct and Fay-Herriot estimates of HIV prevalence in South African districts in 2012, from Gutreuter *et al.* (2019).

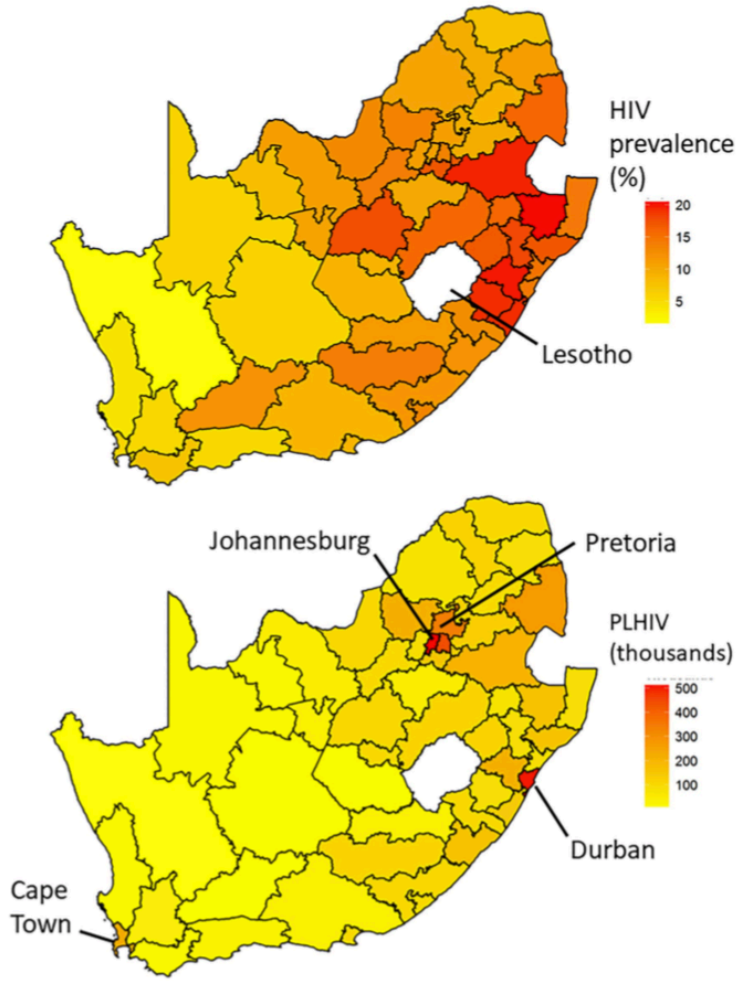


Figure 3: Estimates of HIV prevalence and people living with HIV in South African districts in 2012, from Gutreuter *et al.* (2019).

Discrete Spatio-Temporal Model

- We now move to the space-time setting.
- Let $\widehat{p}_{it}$ be the design-based estimate of a prevalence in area i and period t.
- Take logit of direct estimates $\widehat{p}_{it}$, with appropriate design-based estimator of the variance $\widehat{V}_{it}$, and model as in Mercer *et al.* (2015):

$$\begin{aligned} \text{logit}(\widehat{p}_{it}) &\sim \text{N}(\theta_{it}, \widehat{V}_{it}) \\ \theta_{it} &= \alpha + \underbrace{\phi_t}_{\text{Temporal Smooth}} + \underbrace{\omega_t}_{\text{Temporal Noise}} + \underbrace{e_i}_{\text{Spatial Noise}} + \underbrace{S_i}_{\text{Spatial Smooth}} + \underbrace{\delta_{it}}_{\text{Interaction}} \end{aligned}$$

- Alleviates small sample size problems via temporal, spatial and space-time smoothing.
- Interaction terms are as described by Knorr-Held (2000) and discussed in the Bayes/spatial smoothing lecture.

Space-Time Modeling of U5MR using a Discrete Hazards Model

Motivation

- Aim of this work: In many developing world countries, vital registration is not carried out, so that births and deaths go unreported.
- We aim to provide reliable U5MR estimates at the subnational level, to aid with policy interventions and to assess progress towards health targets. We use data from Demographic Health Surveys (DHS).
- DHS Program: Typically stratified cluster sampling to collect information on population, health, HIV and nutrition; more than 300 surveys carried out in over 90 countries, beginning in 1984.
- The Problem: Data become sparse as we proceed to finer spatial levels.
- The Approach: Leverage space-time similarity to construct a Bayesian smoothing model.

Demographic Health Surveys

- The DHS use stratified (urban/rural, region), 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 are generally high; 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.

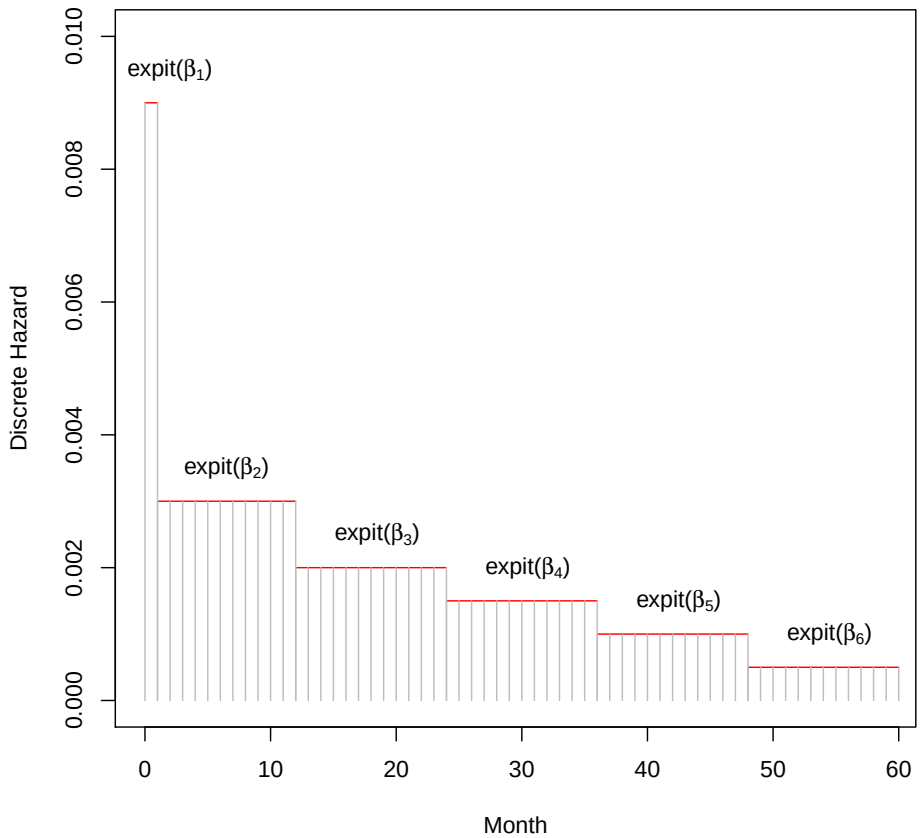
Discrete Survival Model

- We know that infant mortality varies greatly over the first 5 years of life and two possible approaches to modeling how mortality varies with age are:
  - A continuous function of age, via a parametric model (e.g., Weibull, gamma).
  - A discrete function of age, which involves splitting age into intervals.
- For flexibility, we follow the latter route and assume a discrete survival model, with six discrete hazards (probabilities of dying in a particular interval, given survival to the start of the interval) for each of the age bands: $[0, 1)$, $[1, 12)$, $[12, 24)$, $[24, 36)$, $[36, 48)$, $[48, 60]$
- The first category corresponds to neonatal mortality.

Discrete Hazards Model

- We assume a discrete hazard model, with six hazards for each of the age bands: [0,1), [1,12), [12,24), [24,36), [36,48), [48,60].
- In demography speak, ${}_nq_x$ is the risk of death between months x and $x+n$, given survival until x.
- For area i and period t, survival for five years is:

$$1 - {}_{60}q_0^{it} = \prod_{x=0}^{59}(1 - {}_1q_x^{it})$$



Discrete Hazards Model

The estimand of interest in area i and time period t is:

$$
\begin{aligned}
\text{U5MR}^{it} &= {}_{60}q_0{}^{it} \\
&= 1 - \prod_{x=0}^{59}(1 - {}_1q_x{}^{it}) \\
&= 1 - \underbrace{\left[\frac{1}{1+\exp(\beta_1^{it})}\right] \times \left[\frac{1}{1+\exp(\beta_2^{it})}\right]^{11} \times \cdots \times \left[\frac{1}{1+\exp(\beta_6^{it})}\right]^{12}}_{\text{1+11+12+12+12+12 = 60 terms}}
\end{aligned}
$$

- **Design-based inference:** To acknowledge the complex designs weighted logistic regression model (Binder, 1983) is used to estimate the β's, with a standard error based on the design.
- **Model-based inference:** Product of Bernoulli's for each child, with logistic regression model including stratification fixed effects and random effects.

Smoothed Direct Estimation

The smoothed direct model is implemented in the `R` package `SUMMER` (Martin *et al.*, 2018):

- A design object is created in the `survey` package, and direct estimates formed.
- The space-time model is fit using `INLA`.
- It is computationally inexpensive, producing country-specific estimates (for generic indicators or for U5MR) in seconds.

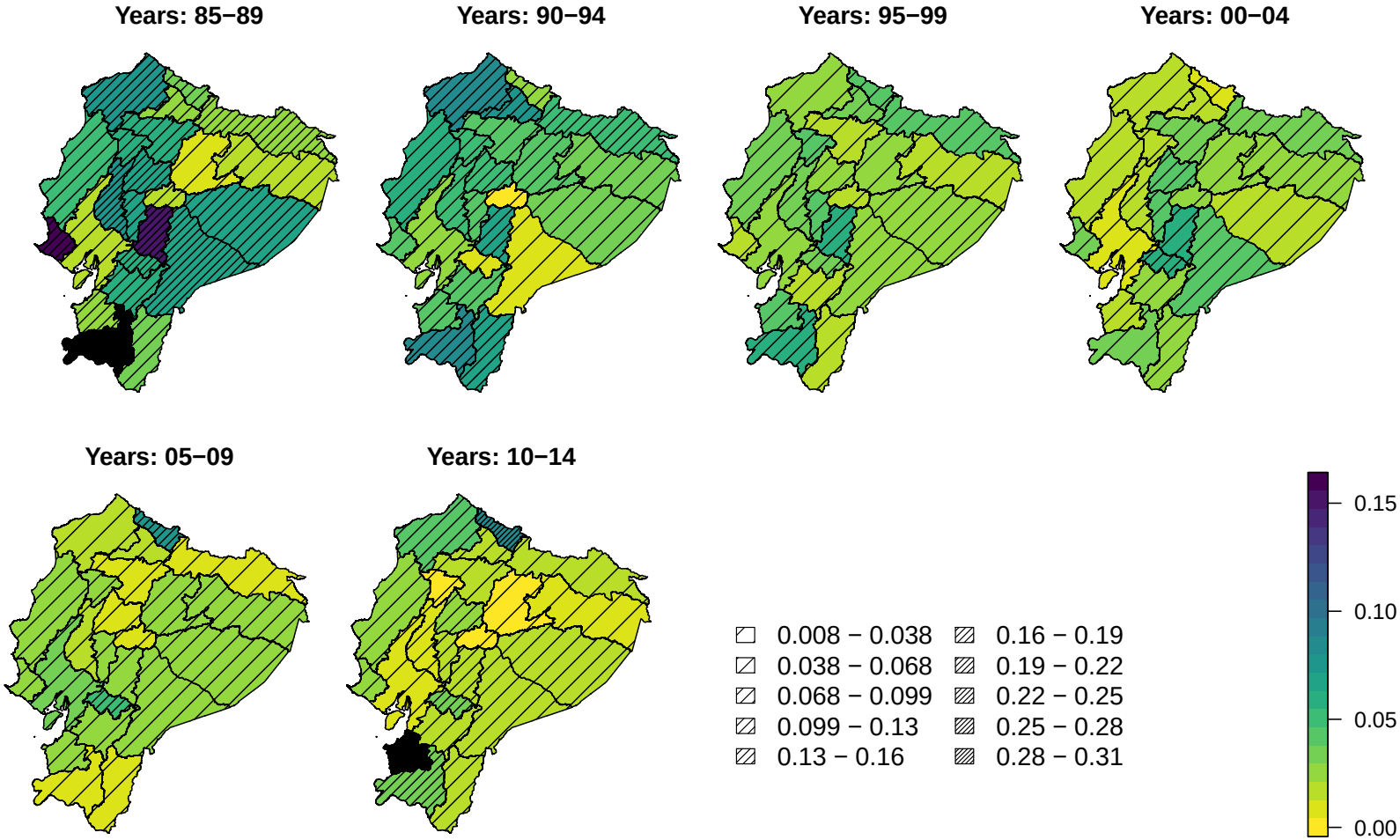


Figure 4: Five-year period weighted estimates (from discrete survival model) of U5MR in Ecuador, with uncertainty indicated by density of hatching; more hatching → more uncertainty, with the latter measured through width of 95% uncertainty interval.

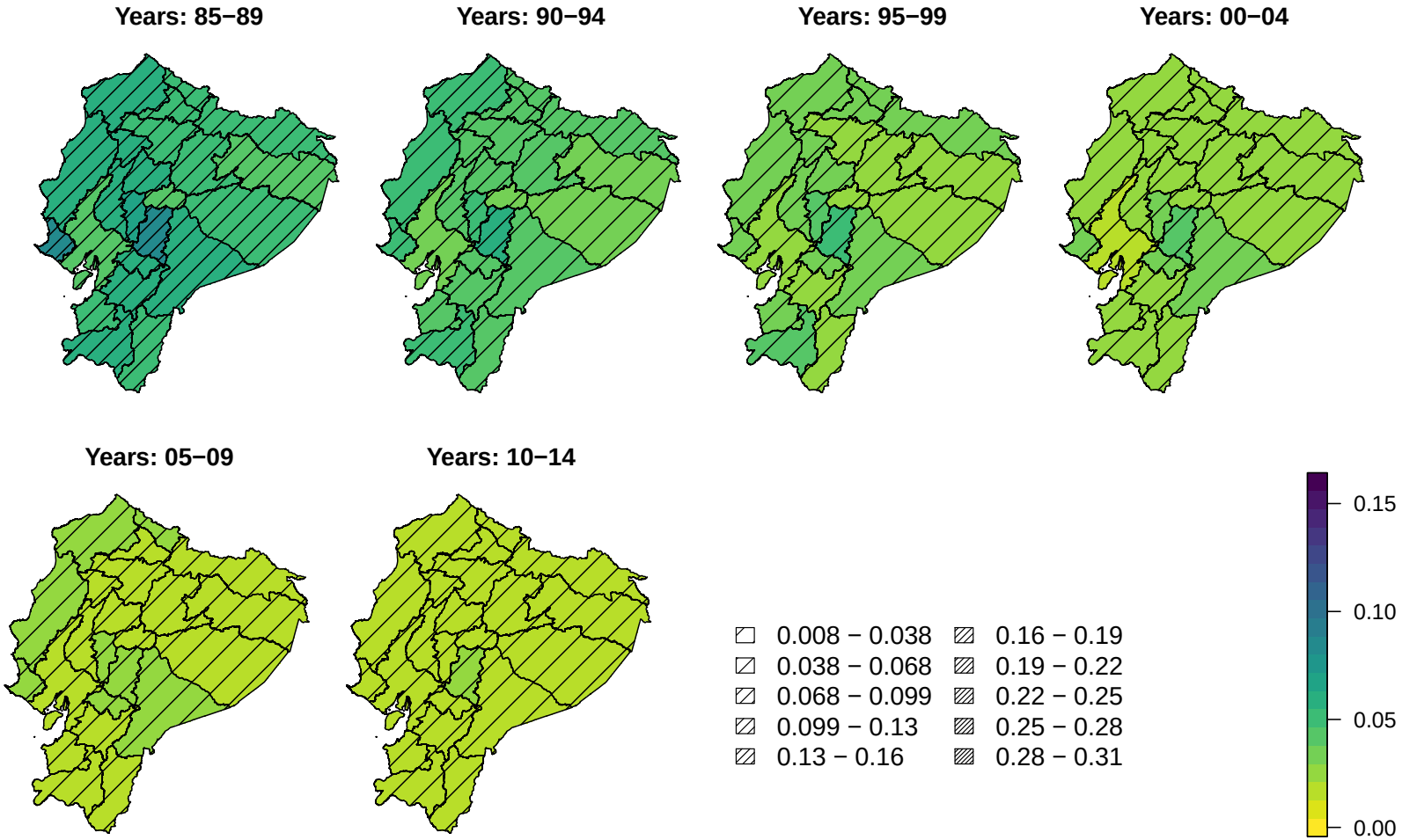


Figure 5: **Smoothed estimates** (from a discrete survival model) of U5MR in Ecuador, with uncertainty indicated by density of hatching; more hatching → more uncertainty, with the latter measured though width of 95% uncertainty interval.

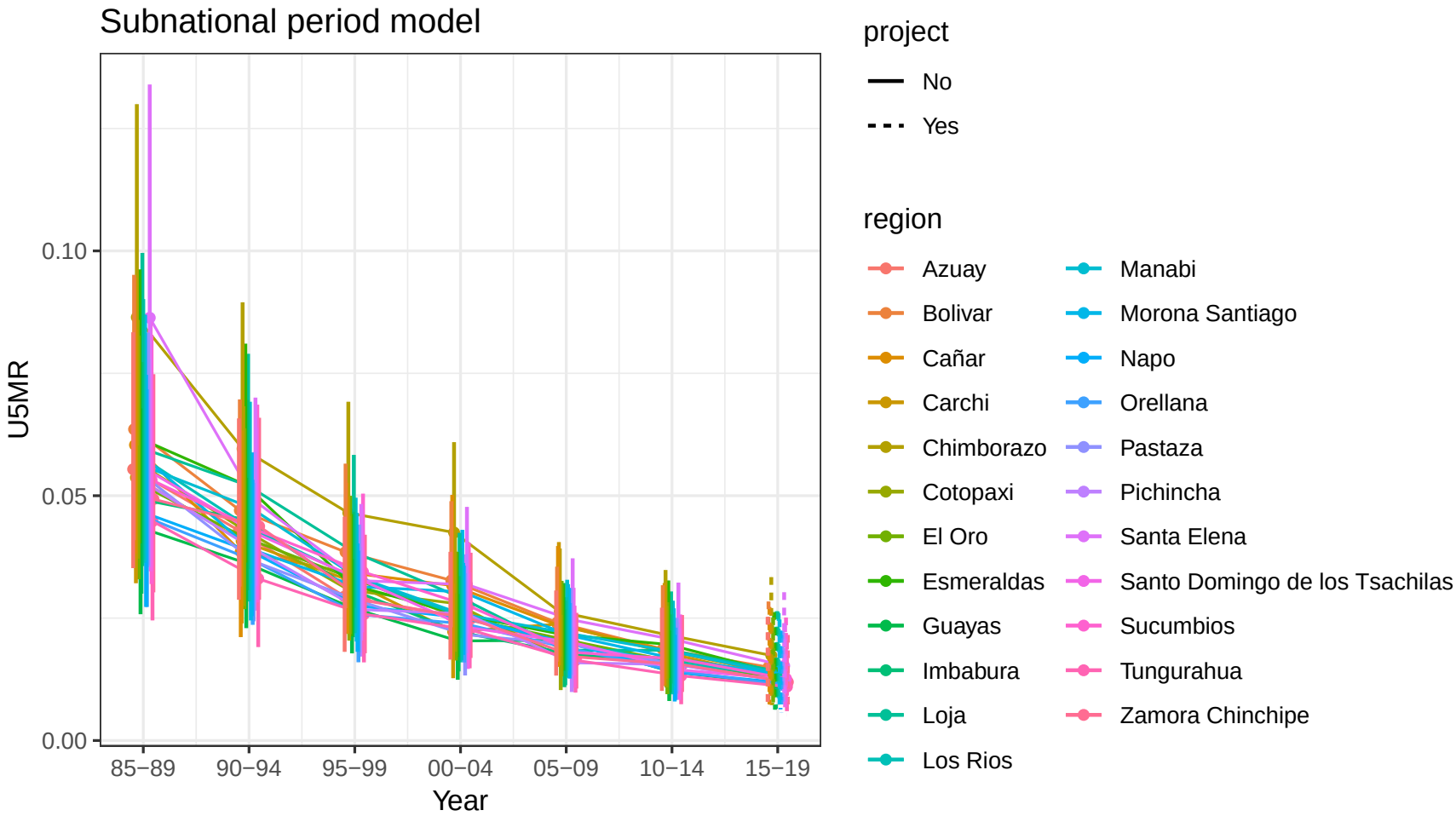


Figure 6: Five-yearly smoothed estimates (from a discrete survival model) of U5MR in Ecuador, by province, with 95% uncertainty intervals.

Smoothed Direct Model at Scale (Li *et al.*, 2019)

- The smoothed direct model has been used for 35 African countries to estimate U5MR in Admin-1 regions, by year.
- Data enter at the 5-year level (to give stable variances), but the RWs are defined on the 1-year scale.
- Data:
  - 121 DHS in 35 countries
  - 1.2 million children
  - 192 million child-months
- Takes around 2.5 hours to obtain estimates for all countries – separate models for each country.
- The smoothed direct model is very reliable for examining Admin1 subnational variation, but the direct estimates are often unreliable for Admin2 estimation.
- Hence, we shortly describe a cluster-level model for this endeavor.

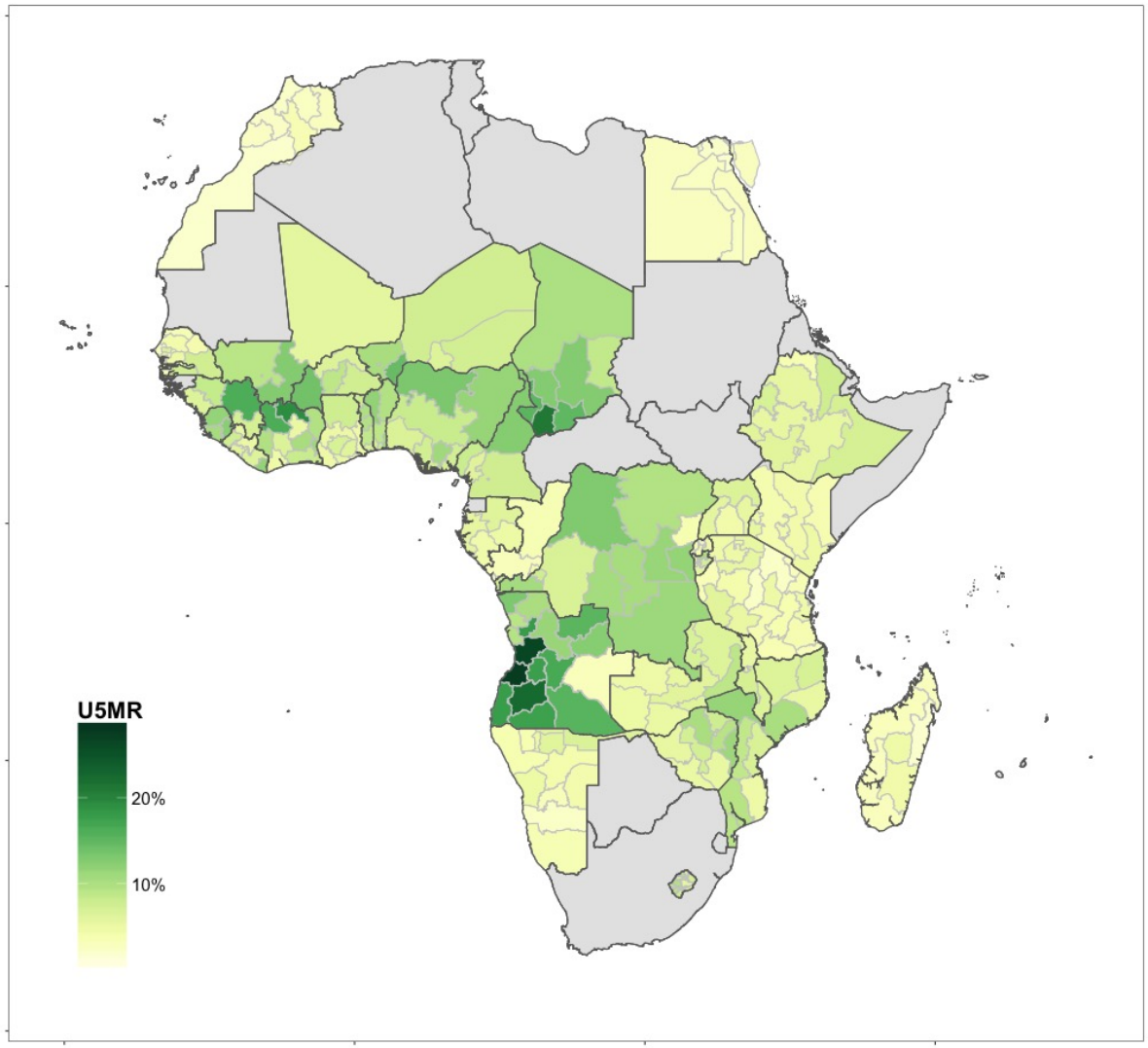


Figure 7: Predictions of U5MR for 2015, in 35 countries of Africa.

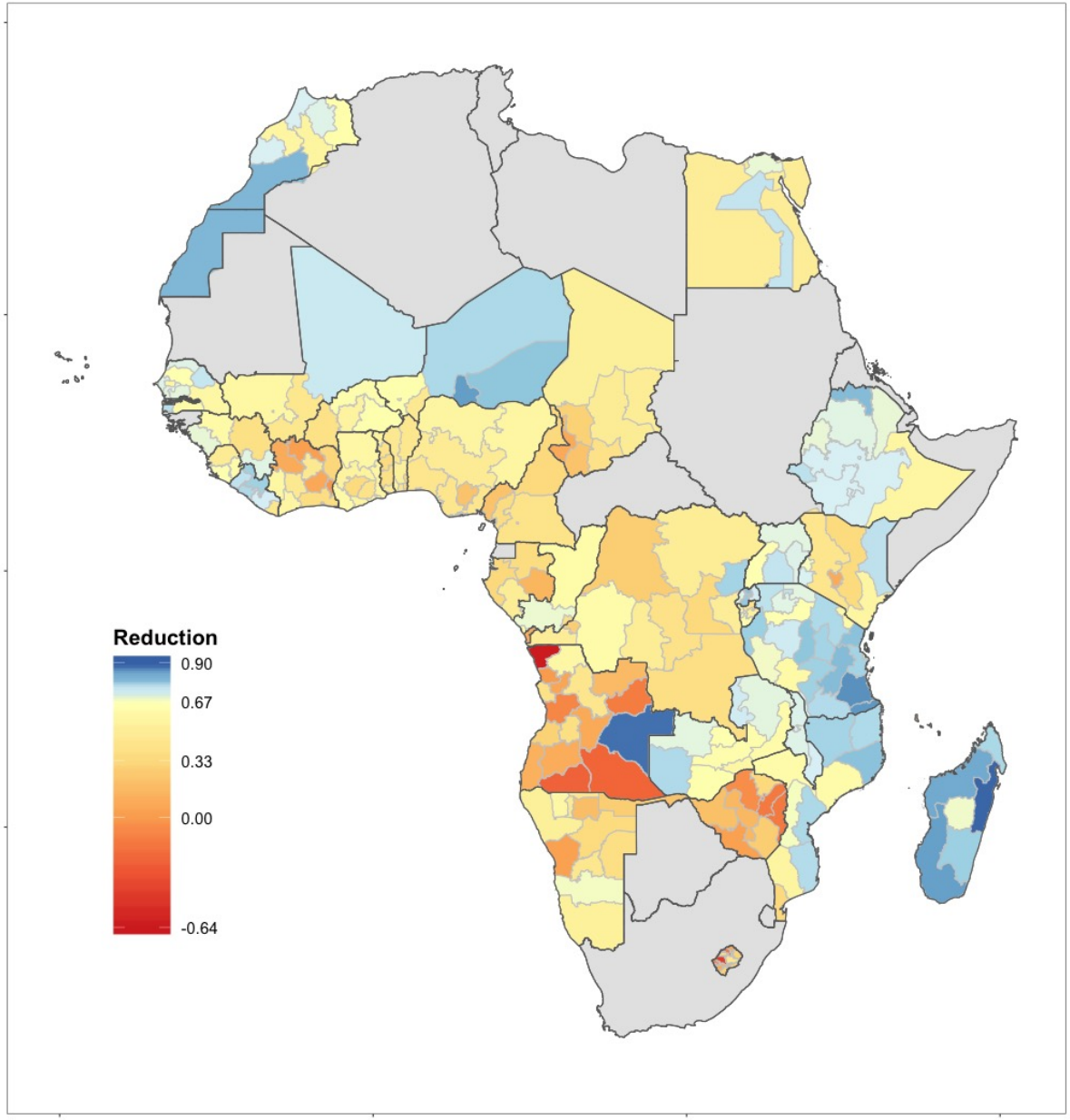


Figure 8: Percent reduction from 1990 to 2015, in 35 countries of Africa.

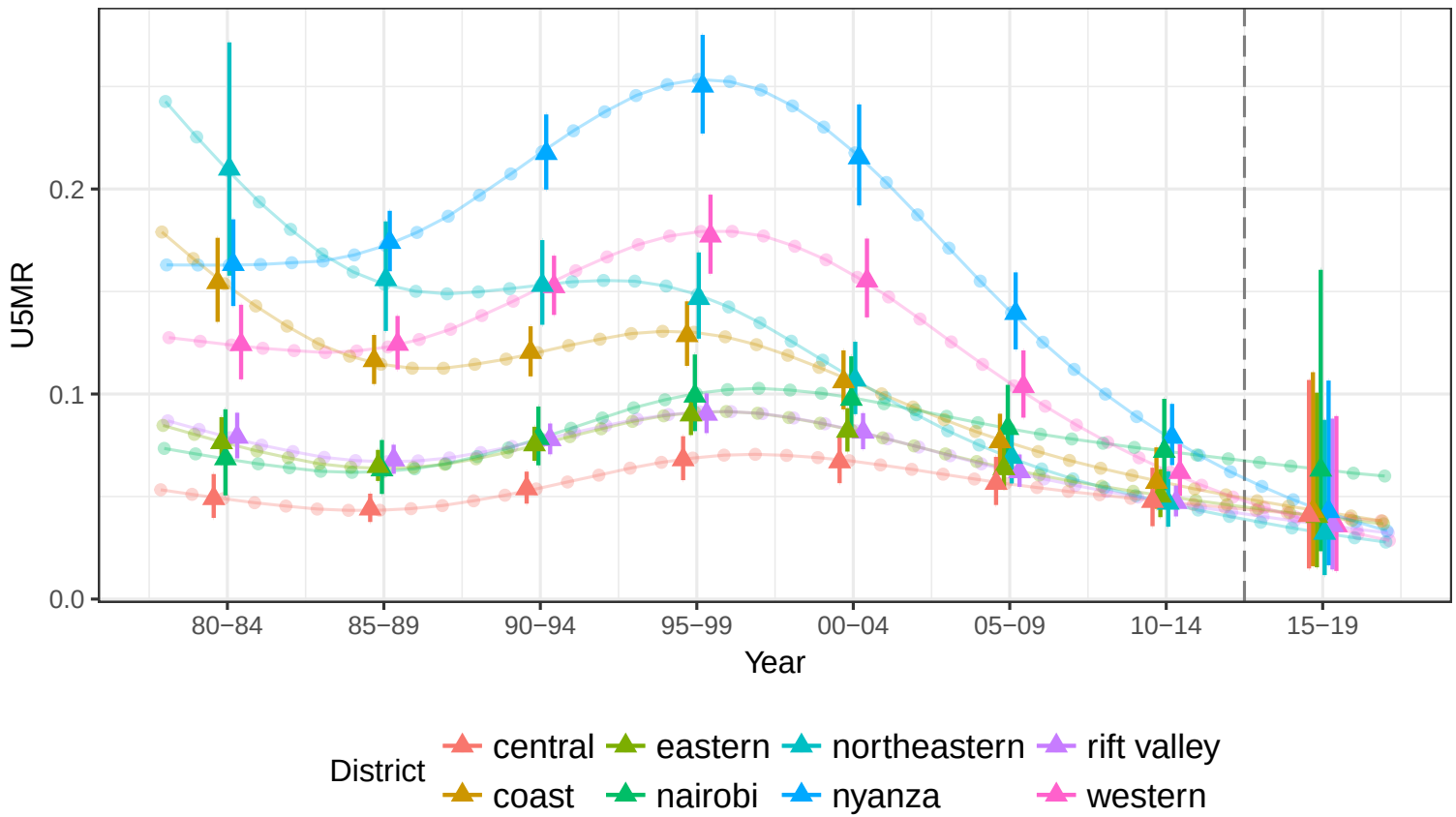


Figure 9: Posterior median estimates for Kenya districts.

Cluster-Level Modeling

Notation

- Initially we describe the model in space only.
- Suppose m_i clusters are sampled within area i, $i = 1, \ldots, n$ of a study area.
- Let $\boldsymbol{s}_c$ represent the geographical location of cluster $c$with $c = 1, \ldots, m$ so that $m = \sum_i m_i$ is the total number of clusters.
- As a concrete example let us take neonatal deaths as the outcome so that the number of deaths and the number of births are denoted $Y(\boldsymbol{s}_c)$ and $n(\boldsymbol{s}_c)$, respectively.

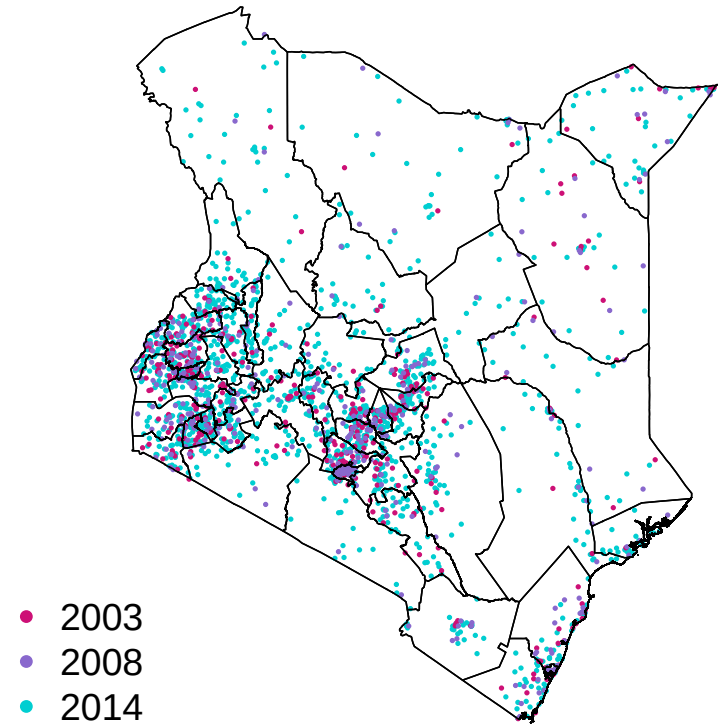


Figure 10: Cluster locations in three Kenya DHS, with county boundaries.

Beta-Binomial Cluster-Level Model

- It is common to see overdispersion with spatial health and demographic data.
- One approach to modeling this phenomenon is to assume the cluster-level prevalence $q_c = q(\boldsymbol{s}_c)$ that is producing the survey data we see in cluster c, is drawn from a probability distribution.
- If we were to go back in time and draw another random sample, a different q_c would result.
- The common q_c to all units sampled, induces correlation between the responses.
- Overdispersion can be modeled using a random effects distribution for the prevalence.
- A common approach is to add a cluster-level normal random effect, see for example Diggle and Giorgi (2019).

Beta-Binomial Cluster-Level Model

- Here, we suppose the cluster level variability is described by the random effects distribution:

$$q_c | a_c, b_c \sim \text{Beta}(a_c, b_c),$$

with $a_c = dp_c$, $b_c = d(1 - p_c)$ so that $d = a_c + b_c$, and

$$\begin{aligned} p_c = \text{E}[q_c] &= \frac{a_c}{d} \\ \text{var}(q_c) &= \frac{p_c(1 - p_c)}{d + 1}. \end{aligned}$$

- The overdispersion is described by the scale parameter d.
- The intraclass correlation coefficient is the correlation between two binary outcomes in the same cluster and corresponds to $1/(d + 1)$.
- The parameters a_c and b_c are not the most intuitive, and it is useful to instead think about the two free parameters as being the mean p_c and the scale d.

Beta-Binomial Cluster-Level Model

- The sampling model corresponds to

$$\begin{aligned} Y_c|q_c &\sim \text{Binomial}(n_c, q_c) \\ q_c|a_c, b_c &\sim \text{Beta}(a_c, b_c) \end{aligned}$$

  which can be integrated over q_c to give the marginal distribution:

$$\Pr(Y_c|p_c, d) = \int_{q_c} \underbrace{\Pr(Y_c|n_c, q_c)}_{\text{Binomial}(n_c, q_c)} \times \underbrace{p(q_c|p_c, d)}_{\text{Beta}(a_c, b_c)} \, dq_c,$$

  with $p_c = a_c/(a_c + b_c)$ and $d = 1/(a_c + b_c)$.
- Turning the handle,

$$Y_c|p_c, d \sim \text{Beta-Binomial}(n_c, p_c, d),$$

  with

$$\begin{aligned} \text{E}[Y_c|p_c, d] &= n_c p_c \\ \text{var}(Y_c|p_c, d) &= n_c p_c(1 - p_c) \times \frac{n_c + d}{1 + d}, \end{aligned}$$

  so we have overdispersion for $d > 0$.

Beta-Binomial Cluster-Level Model

- We still need to specify a form for the mean, and a logistic model is natural.
- Under the discrete spatial model,

$$p_c = p(\boldsymbol{s}_c) = \text{expit}(\alpha + z_c\gamma + e_{i[s_c]} + S_{i[s_c]}),$$

  where the notation $i[s_c]$ here should be read as "the area i which contains the cluster at location s_c".
- The constituent terms are:
  - $p(\boldsymbol{s}_c)$ is the prevalence at location $\boldsymbol{s}_c$,
  - z_c is the strata within which cluster c lies (with $z_c = 0/1$ for urban/rural),
  - $\exp(\gamma)$ is the associated odds ratio,
  - e_i is an IID area-level error term and S_i is a spatial ICAR random effect.
- Note that this model is the binomial version of the two-fold nested error regression model, as discussed in Section 4.5.2 of Rao and Molina (2015).

Aggregation

- With this simple spatial form, the modeled area prevalence is

$$p_i = r_i \times \underbrace{\text{expit}(\alpha + S_i + e_i)}_{\text{Prevalence for Urban}} + (1 - r_i) \times \underbrace{\text{expit}(\alpha + \gamma + S_i + e_i)}_{\text{Prevalence for Rural}},$$

  where
  - r_i is the proportion of the area that is urban, and
  - $1 - r_i$ the proportion that is rural.
- The original sampling frame that contains the proportions of urban/rural, is unavailable, though some information is typically available in the DHS reports.
- The proportions r_i can be obtained by thresholding population density surfaces.

Stochastic Partial Differential Equations (SPDEs)

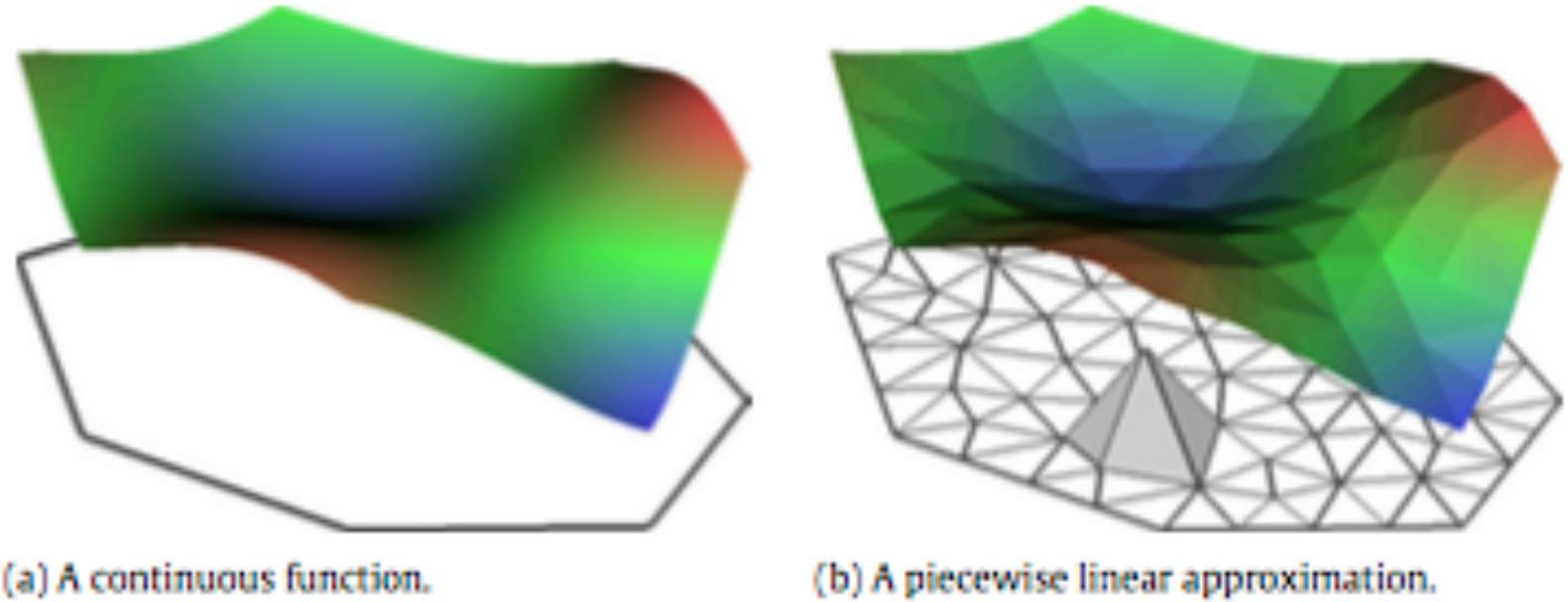
- We briefly describe a cluster-level model with a continuous spatial field.
- The sampling model is $Y_c|p_c \sim \text{Binomial}(n_c, p_c)$, with

$$p_c = p(\boldsymbol{s}_c) = \text{expit}(\alpha + z_c\gamma + S(\boldsymbol{s}_c) + \epsilon_c),$$

  where
  - $p(\boldsymbol{s}_c)$ is the prevalence at location $\boldsymbol{s}_c$,
  - z_c is the strata within which cluster c lies (with $z_c = 0/1$ for urban/rural),
  - $\exp(\gamma)$ is the associated odds ratio,
  - $\epsilon_c \sim \text{N}(0, \sigma^2_\epsilon)$ is the so called nugget which represents short scale variation and/or "measurement error"
  - $S(\boldsymbol{s})$ is a spatial Gaussian process (GP) random effect – in the results that follow we implement using the SPDE approach (Lindgren *et al.*, 2011).
- Aggregation requires the population density $p(\boldsymbol{s})$:

$$p_i = \int_{A_i} p(\boldsymbol{s})d(\boldsymbol{s})\, d\boldsymbol{s}.$$

SPDE Approximation



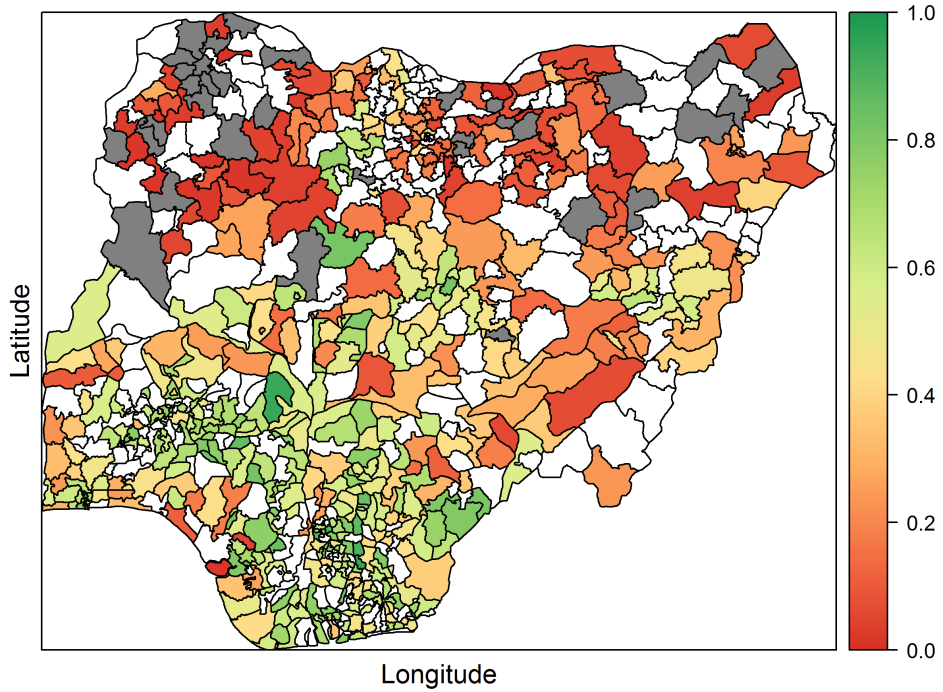
(a) A continuous function. (b) A piecewise linear approximation.

Fig. 2. Piecewise linear approximation of a function over a triangulated mesh.

Figure 11: GMRF representation of a Markovian GRF, via triangulation, from Simpson *et al.* (2012)

Application to Vaccination Prevalence in Nigeria

- We examine cluster-level spatial smoothing models in the context of estimating measles vaccination rates in Nigeria in 2013, using the 2013 Nigerian DHS.
- In Nigeria, the Admin2 areas correspond to Local Government Areas (LGAs) and there are 774 in total – with such a large number there are many LGAs with little/no data.
- There are no clusters in 255 LGAs (shown in white below).



Application to Vaccination Prevalence in Nigeria

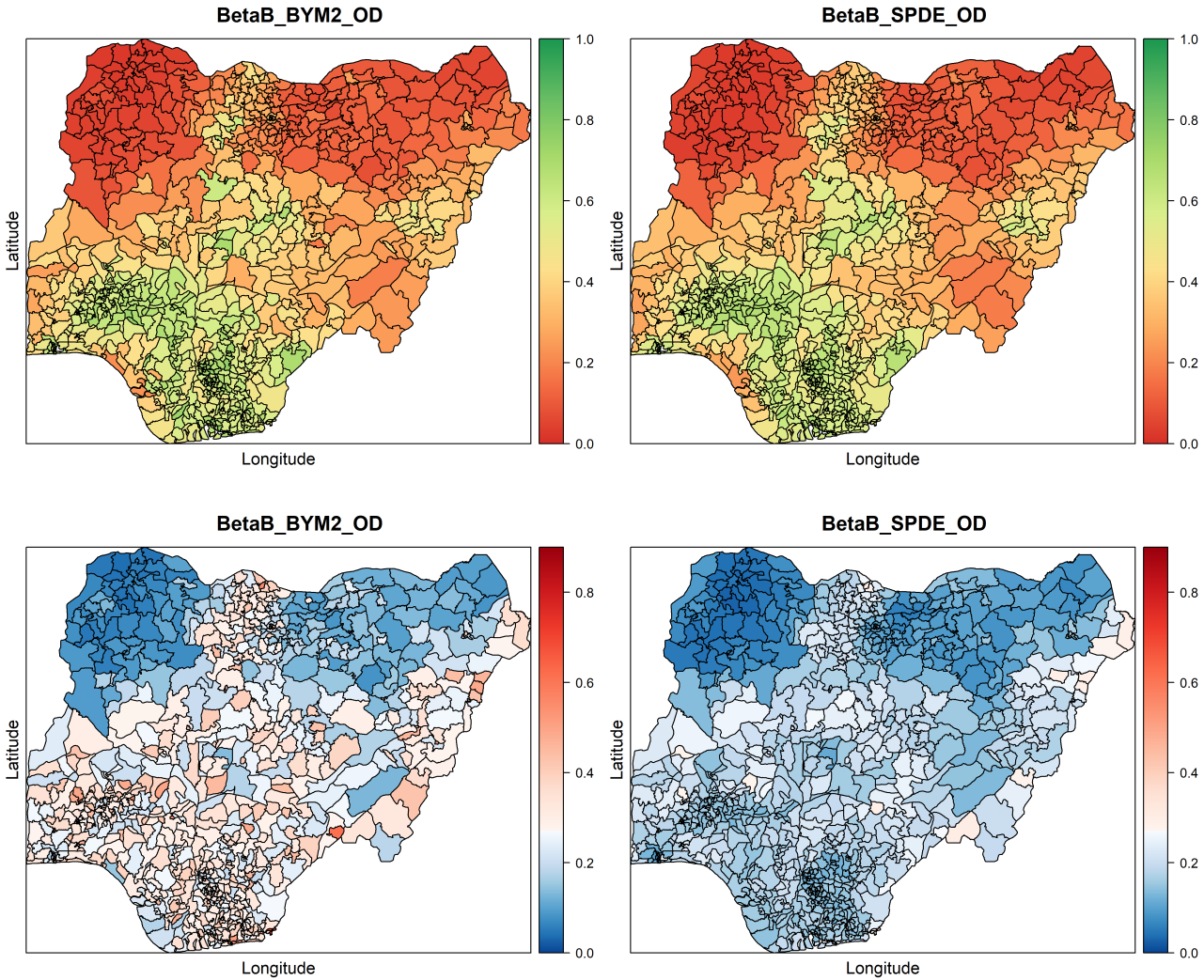
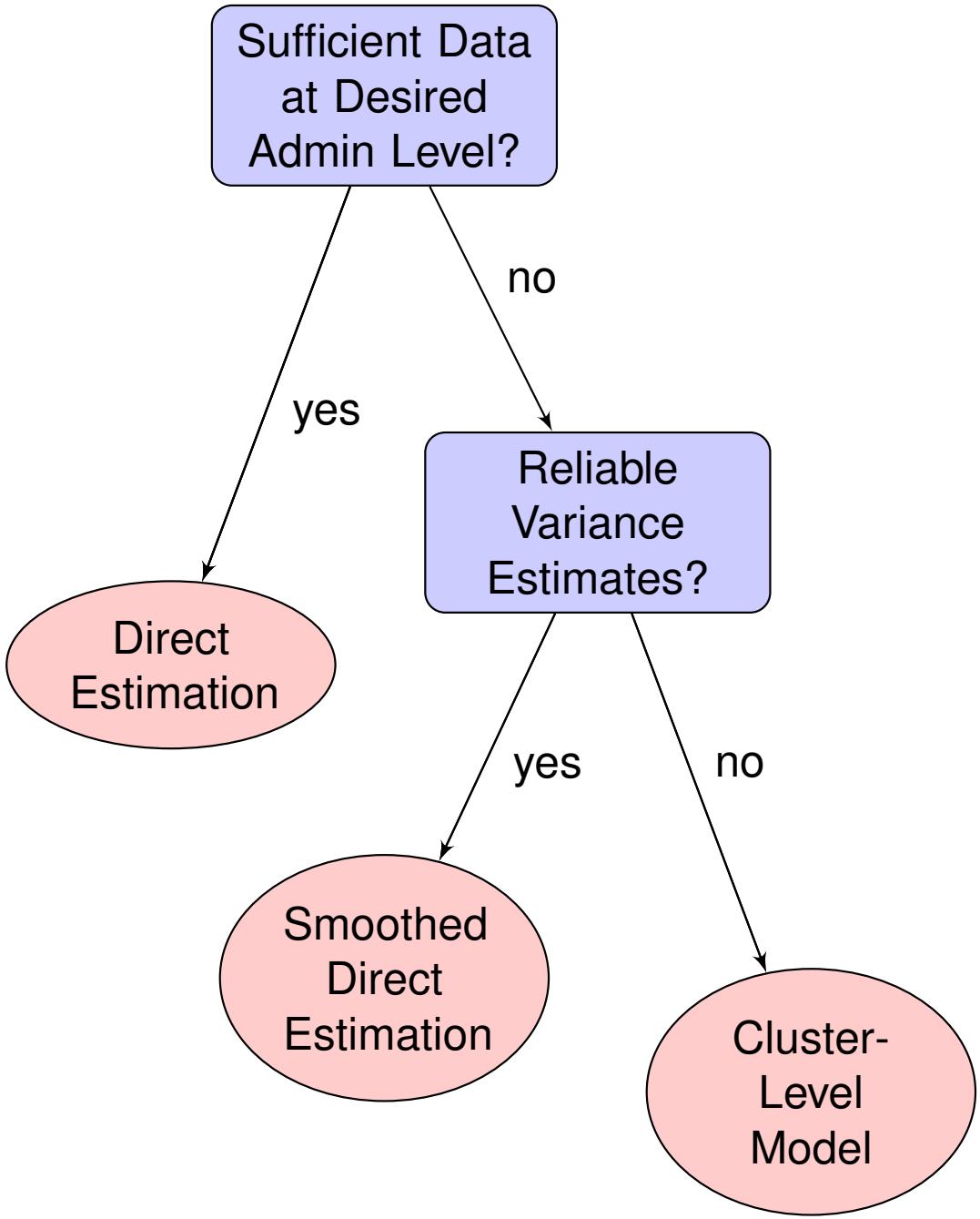


Figure 13: Top row: posterior estimates under BYM (left) and SPDE (right) models. Bottom right: width of 90% intervals for BYM (left) and SPDE (right).

Recommended Methods for Routine Work



Discussion

Discussion

- The smoothed direct model builds on the strengths of direct (weighted) estimates and spatial smoothing models.
- In the limit, as we obtain larger data in an area, the weighted estimates will dominate, which is exactly what we want!
- If insufficient samples in areas, then estimated variance is unacceptably large (or undefined), and then we need to resort to the cluster-level models:
  - Discrete spatial models are easier to fit, and aggregation more straightforward.
  - Continuous spatial models are more challenging to fit, and aggregation more challenging.
- Model checking techniques are still in their infancy in the SAE context.
- Prevalence mapping is still in its infancy, and currently no agreed upon “best” approach.

Acknowledgments

This lecture series was supported by the Hewlett Foundation and the International Union for the Scientific Study of Population (IUSSP).

The research reported in this series has grown out of a longstanding collaboration between Jon Wakefield, Zehang Richard Li and Sam Clark.

Many other people have contributed, however, for full details and for links to other aspects of this work, check out:

http://faculty.washington.edu/jonno/space-station.html

References

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

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

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

Gutreuter, S., Igumbor, E., Wabiri, N., Desai, M., and Durand, L. (2019). Improving estimates of district HIV prevalence and burden in South Africa using small area estimation techniques. *PLoS One*, **14**, e0212445.

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

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

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

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

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

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

Pfeffermann, D. (2013). New important developments in small area estimation. *Statistical Science*, **28**, 40–68.

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

Simpson, D., Lindgren, F., and Rue, H. (2012). Think continuous: Markovian Gaussian models in spatial statistics. *Spatial Statistics*, **1**, 16–29.