2022 SISCER Module 2 Small Area Estimation:
Lecture 2: Area-Level Modeling

Jon Wakefield

Departments of Statistics and Biostatistics
University of Washington

Outline

Motivation for Mixed Models and Smoothing

Motivation

In a SAE context, we are often faced with situations in which the data are sparse in space, which leads to great uncertainty in calculated estimates.

Mixed models[1] are designed to alleviate this problem, by modeling the totality of data from all areas, in order to leverage similarities in the data – these are examples of indirect estimates because they use response data from areas other than the area for which an estimate is required.

The key element is coupling the different areas, by assuming this parameters in all areas are linked through a common probability distribution.

In this lecture we describe mixed models for normal data (via linear mixed models) with spatial smoothing models being included.

[1] also known as random effects or hierarchical models

Model-based approaches

Indirect methods provide a link, often with an implicit model, between different areas; we describe explicit models that provide such a link.

The models we describe are mixed-effects models which aim to accurately describe between domain (area) differences.

Such models offer several advantages:

- Models can be tuned to the application, building on the existing theory and practical experience of mixed models, including non-linear models, such as logistic mixed models.
- Domain-(Area)-specific measures of uncertainty are produced.
- One can attempt to check assumptions using diagnostics.
- A variety of area-specific random effects models, including spatial versions, are available.

Model-based approaches

The use of explicit models has not been carried out greatly in survey sampling, where design-based inference is historically the norm.

Models can be specified at the level of the area or the (observation) unit.

Disadvantages of mixed models:

- How to incorporate the design weights/acknowledge the design?
- In area-level (Fay-Herriot) models, the design is explicitly considered while for unit-level models we need terms in the model.
- It is often difficult to check modeling assumptions.
- Computation can be demanding, though this is improving.

Motivating Examples: Simulated Normal Data

We consider simulated normal data: using the King County health reporting areas (HRAs) geography.

These data were simulated with non-constant mean across HRAs (areas).

Nominally, the outcome will be labeled as weight.

Motivating Example: Simulated Normal Data

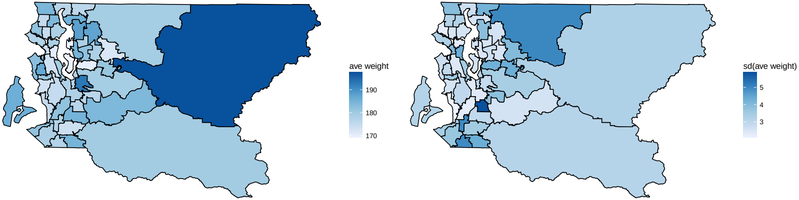

Figure 1: Sample mean weights (left) and standard errors of mean weights (right).

Linear Mixed Effects Models (LMEMs)

Smoothing Models

Instability of estimates has lead to methods being developed to impose smoothness on the underlying parameters using mixed effects models that use the data from the totality of areas to provide more reliable estimates in each of the constituent areas.

Overview of Models:

- Basic Linear Model: No smoothing.
- Linear Mixed Models:
  - Normal likelihood. Random effects with no spatial structure (known as IID[2]).
  - Normal Likelihood, Two sets of random effects, one set with no spatial structure, one set with spatial structure.
- Covariates may be added to each of these in order to smooth over covariate space.
- Estimation in these models is a separate issue.

[2] Independent and Identically Distributed

Linear Mixed Effects Model

Let Y_{ik} be the weight of the k-th sampled individual (unit) in area i.

Three possible models:

1. No between-area variability:

$$Y_{ik} = \underbrace{\beta_0}_{\text{Common Mean}} + \epsilon_{ik},$$

   with $\epsilon_{ik} \sim \mathrm{N}(0, \sigma_\epsilon^2)$. (Most basic synthetic estimates).

2. Distinct between-area variability:

$$Y_{ik} = \underbrace{\beta_i}_{\text{Mean of Area } i} + \epsilon_{ik},$$

   with $\epsilon_{ik} \sim \mathrm{N}(0, \sigma_\epsilon^2)$. Known as a fixed effects model. Note: no way to link different areas. (Direct estimates).

3. Linked between-area variability:

$$Y_{ik} = \underbrace{\beta_0 + \delta_i}_{\text{Mean of Area } i} + \epsilon_{ik},$$

   with $\delta_i \sim \mathrm{N}(0, \sigma_\delta^2)$, $\epsilon_{ik} \sim \mathrm{N}(0, \sigma_\epsilon^2)$. Known as a mixed effects model. (Indirect estimates).

Basic Linear Mixed Effects Model

We will concentrate on the linked between-area mixed effects model:

$$Y_{ik} = \underbrace{\beta_0 + \delta_i}_{\text{Mean of Area } i} + \epsilon_{ik},$$

with $\delta_i \sim \text{N}(0, \sigma^2_\delta)$ – these are the area-specific deviations (the random effects) from the overall level β_0 – and $\epsilon_{ik} \sim \text{N}(0, \sigma^2_\epsilon)$, is the measurement error.

This model is also known as a Linear Mixed Effects Model (LMEM).

In this model, the totality of data are used to inform on the overall level β_0 and between-area variability σ^2_δ.

The unknown parameters are:

Overall mean	β_0
Between-Area Variance	σ^2_δ
Measurement Error Variance	σ^2_ϵ
Random Effects	$\delta_1, \ldots, \delta_n$

Maximum Likelihood Estimation

The MLEs are the values of the parameters that maximize the probability of the observed data, under an assumed probability model.

For a LMEM the likelihood to be maximized is,

$$L(\beta_0, \sigma_\delta^2, \sigma_\epsilon^2) = \prod_{i=1}^{n} \int_{\delta_i} p(\boldsymbol{y}_i | \beta_0, \delta_i, \sigma_\epsilon^2) \times p(\delta_i | \sigma_\delta^2) \, d\delta_i.$$

If a likelihood approach is taken, the random effect estimates $\widehat{\delta}_i$, are obtained as the conditional means $\mathrm{E}[\delta_i | \boldsymbol{y}, \beta_0, \sigma_\epsilon^2, \sigma_\delta^2]$.

These are known as the best linear unbiased predictors (BLUPs).

Known as estimated BLUPs (EBLUPs) when β_0 and variance components $\sigma_\epsilon^2, \sigma_\delta^2$ are estimated.

Technical Note: Restricted ML (REML) is preferred for estimation of variances – accounts for estimation of $\boldsymbol{\beta}$.

Estimation in the Linear Mixed Model

In general, there are no closed-form (i.e., explicit) forms for the estimates of the parameters.

Suppose, for simplicity, that β_0, σ_δ^2 and σ_ϵ^2 are known; this allows some insight into inference.

Estimation in the Linear Mixed Model

The posterior mean of the random effect (area-specific adjustment) is:

$$
\begin{aligned}
\widehat{\delta}_i &= \mathrm{E}[\delta_i | y_i, \beta_0, \sigma^2_\epsilon, \sigma^2_\delta] \\
&= \frac{n_i \sigma^2_\delta}{\sigma^2_\epsilon + n_i \sigma^2_\delta} (\overline{y}_i - \beta_0) \\
&= q_i (\overline{y}_i - \beta_0)
\end{aligned}
$$

where

$$
q_i = \frac{n_i \sigma^2_\delta}{\sigma^2_\epsilon + n_i \sigma^2_\delta} \leq 1
$$

and is small (so more shrinkage) if:

- n_i is small (not much data in the area), or
- σ^2_δ is small (between-area variability is small), or
- σ^2_ϵ is large (within-area variability is large).

$\widehat{\delta}_i$ is the BLUP.

Inference for Random Effcets

From a frequentist mixed-model perspective:

- The BLUPs are unbiased in the sense that,

$$\mathrm{E}[\widehat{\delta}_i - \delta_i] = 0,$$

  where the expectation is over $\widehat{\delta}_i$ and δ_i since both are viewed as random.
- Uncertainty is measured by $\mathrm{var}(\widehat{\delta}_i - \delta_i)$ which is equal to the mean squared error (MSE), $\mathrm{MSE}(\widehat{\delta}_i)$.
- The MSE can also be used for construction of asymptotic confidence intervals:

$$\widehat{\delta}_i - \delta_i \sim \mathrm{N}(0, \underbrace{\mathrm{MSE}(\widehat{\delta}_i)}_{\mathrm{var}(\widehat{\delta}_i - \delta_i)}).$$

- The MSE can be estimated analytically, or via the bootstrap (Rao and Molina, 2015).
- Note that the frequentist coverage is obtained over repeated sampling of $\boldsymbol{y}$ and $\boldsymbol{\delta}$, which is not what is practically useful, see Burris and Hoff (2020) for more discussion.

Predicting the Population Total and Mean

Let S_i and R_i denote, respectively, the set of indices of the sampled and unsampled individuals in area i, with n_i sampled individuals.

Let $T_i = \sum_{k=1}^{N_i} y_{ik}$ be the total for the population in area i, where N_i is the population size.

The average for the population in area i is

$$\begin{aligned}
\overline{Y}_i &= \frac{T_i}{N_i} \\
&= \frac{\sum_{k=1}^{N_i} y_{ik}}{N_i} \\
&= \frac{\sum_{k\in S_i} y_{ik} + \sum_{k\in R_i} y_{ik}}{N_i} \\
&= \underbrace{\frac{\sum_{k\in S_i} y_{ik}}{n_i}}_{\text{Mean of Sampled}} \times \frac{n_i}{N_i} + \underbrace{\frac{\sum_{k\in R_i} y_{ik}}{N_i - n_i}}_{\text{Mean of Unsampled}} \times \frac{N_i - n_i}{N_i}
\end{aligned}$$

Predicting the Population Total and Mean

When we estimate the mean of the unsampled, we are assuming that we have accounted for systematic differences (such as through stratification) of sampled and unsampled individuals.

Suppose now we have fitted the linear mixed effects model and obtained posterior medians $\widehat{\beta}_0$ and $\widehat{\delta}_i$.

The obvious estimate is:

$$\widehat{\overline{Y}}_i = \underbrace{\frac{\sum_{k \in S_i} y_{ik}}{n_i}}_{\text{Mean of Sampled}} \times \frac{n_i}{N_i} + \underbrace{(\widehat{\beta}_0 + \widehat{\delta}_i)}_{\text{Estimated Mean}} \times \frac{N_i - n_i}{N_i}$$

If $N_i \gg n_i$, then the sampled data provide a small fraction of the total population in the area and we can estimate the area mean by

$$\widehat{\overline{Y}}_i = \widehat{\beta}_0 + \widehat{\delta}_i. \quad (1)$$

If an area contains no data, then its mean is assumed to be $\beta_0 + \delta^\star$, where $\delta^\star \sim \text{N}(0, \sigma_\delta^2)$.

Motivating Example: Continuous Outcome

Totals can be similarly estimated:

$$\begin{aligned}\widehat{T}_i &= \sum_{k \in S_i} y_{ik} + \sum_{k \in R_i} y_{ik} \\ &= \underbrace{\sum_{k \in S_i} y_{ik}}_{\text{Total of Sampled}} + \underbrace{(N_i - n_i) \times (\widehat{\beta}_0 + \widehat{\delta}_i)}_{\text{Estimated Total for Unsampled}} .\end{aligned}$$

If $N_i \gg n_i$,

$$\widehat{T}_i \approx N_i \times (\widehat{\beta}_0 + \widehat{\delta}_i).$$

This assumes that population size N_i is known.

Bayesian Inference

Bayesian Inference

Bayesian modeling is convenient for implementing notions of smoothing.

There are two key elements that must be specified:

- The sampling model (likelihood) describes the distribution of the data – this model depends on unknown parameters, that we will denote p.
- The prior distribution expresses beliefs about the parameters p and provides a mechanism by which penalization/smoothing can be imposed.

These elements are probabilistically combined via **Bayes Theorem:**

$$\underbrace{p(p|y)}_{\textbf{Posterior}} \propto \underbrace{L(p)}_{\textbf{Likelihood}} \times \underbrace{\pi(p)}_{\textbf{Prior}}.$$

On the log scale:

$$\underbrace{\log p(p|y)}_{\textbf{Updated Beliefs}} = \underbrace{\log L(p)}_{\textbf{Sampling Model}} + \underbrace{\log \pi(p)}_{\textbf{Penalization}}.$$

Bayesian Inference

- In a Bayesian analysis the complete set of unknowns (parameters) is summarized via the multivariate posterior distribution – one or two dimensional marginal posterior distributions can be vizualised.
- The marginal distribution for each parameter may be summarized via its mean, standard deviation, or quantiles.
- It is common to report the posterior median and a 90% or 95% posterior range for parameters of interest.
- The range that is reported is known as a credible interval.

Bayesian Computation

- The computations required for Bayesian inference (integrals) are often not trivial and may be carried out using a variety of analytical, numerical and simulation based techniques.
- We use the integrated nested Laplace approximation (INLA), introduced by Rue *et al.* (2009).
- `R-INLA` is a package that implements the INLA approach.
- The `SUMMER` package uses INLA for all Bayesian computation.
- Book-length treatments on INLA:
  - Blangiardo and Cameletti (2015) – space-time modeling.
  - Wang *et al.* (2018) – general modeling.
  - Krainski *et al.* (2018) – advanced space-time modeling.

Bayes Example

- Imagine the data model is normal with an unknown mean μ:

$$y_i \mid \mu \sim \mathrm{N}(\mu, \sigma^2),$$

$i = 1, \ldots, n$, with σ^2 assumed known.

- This is equivalent to:

$$\overline{y} \mid \mu \sim \mathrm{N}(\mu, \sigma^2/n),$$

where $\sigma/\sqrt{n}$ is the standard error.

- Suppose a normal prior is appropriate:

$$\mu \sim \mathrm{N}(m, v),$$

so that values of the mean μ that are (relatively) far from m are penalized – v is the smoothing parameter, more/less smoothing if small/big.

- The log posterior is:

$$\underbrace{\log p(\mu \mid y)}_{\text{Updated Beliefs}} = -\underbrace{\frac{n}{2\sigma^2}(\overline{y} - \mu)^2}_{\text{Data Model}} - \underbrace{\frac{1}{2v}(\mu - m)^2}_{\text{Penalization}}.$$

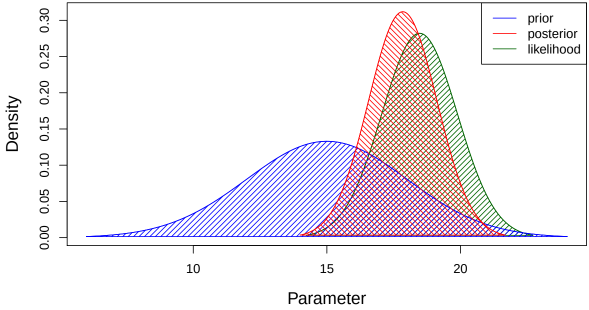

Figure 2: Normal data model with $n = 10$, $\bar{y} = 19.3$ and standard error 1.41. The prior for μ has mean $m =$15 and $v = 3^2$. The posterior for the parameter μ is a compromise between the two sources of information: the posterior mean is 18.5 and the posterior standard deviation is 1.28.

The Bayesian Linear Mixed Model

The LMEM with priors is:

$$
\begin{aligned}
Y_{ik} | \boldsymbol{\beta}, \delta_i, \sigma_\epsilon^2 &\sim_{iid} \mathrm{N}(\beta_0 + \boldsymbol{x}_i^\intercal \boldsymbol{\beta}_1 + \delta_i, \sigma_\epsilon^2) \\
\delta_i | \sigma_\delta^2 &\sim_{iid} \mathrm{N}(0, \sigma_\delta^2) \\
\boldsymbol{\beta}, \sigma_\epsilon^2, \sigma_\delta^2 &\sim \text{Priors}
\end{aligned}
$$

The posterior, given data $\boldsymbol{y}$, is obtained as

$$
\begin{aligned}
p(\boldsymbol{\beta}, \delta_1, \ldots, \delta_n, \sigma_\epsilon^2, \sigma_\delta^2 | \boldsymbol{y}) &= p(\boldsymbol{y} | \boldsymbol{\beta}, \delta_1, \ldots, \delta_n, \sigma_\epsilon^2, \sigma_\delta^2) \\
&\times p(\boldsymbol{\beta}, \delta_1, \ldots, \delta_n, \sigma_\epsilon^2, \sigma_\delta^2) / p(\boldsymbol{y}) \\
&= \prod_{i=1}^{n} p(\boldsymbol{y}_i | \boldsymbol{\beta}, \delta_i, \sigma_\epsilon^2) \times p(\delta_i | \sigma_\delta^2) \\
&\times p(\boldsymbol{\beta}, \sigma_\epsilon^2, \sigma_\delta^2) / p(\boldsymbol{y})
\end{aligned}
$$

Marginal distributions, such as $p(\beta_0 | \boldsymbol{y})$, are obtained by integration.

In the `SUMMER` package, the `INLA` approach is used (Rue *et al.*, 2009).

Simulated Data

Motivating Example: Linear Model

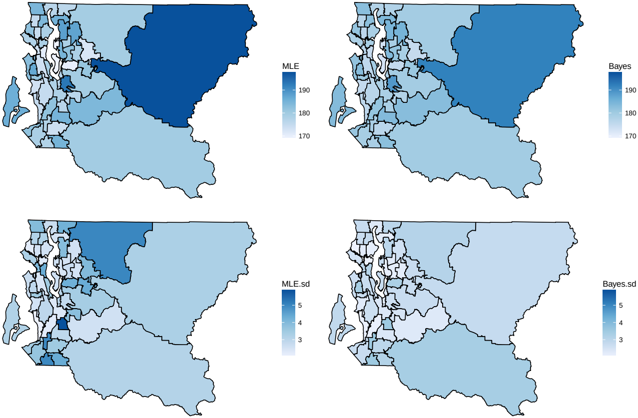

Figure 3: Top row: Estimates of area averages of weight via MLE's (left) and posterior medians (right). Bottom row: Uncertainty of estimates with standard errors (left) and posterior standard deviations (right).

Motivating Example: Linear Model

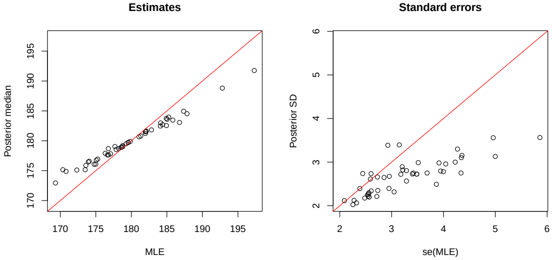

Figure 4: Comparison of area averages: Posterior medians versus MLEs (left). Posterior standard deviations versus standard errors associated with the MLEs (right).

The posterior medians are shrunk from the MLEs towards the overall mean, with the extreme values undergoing the most shrinkage.

In general, the Bayes measures of uncertainty (the posterior standard deviations) are smaller than the standard errors of the MLEs, with the greatest difference occurring for those areas with the large standard errors (which have the smallest sample sizes).

Fay-Herriot Modeling

Fay-Herriot Model

What about accounting for complex sampling?

So far we have been (implicitly) assuming the data are gathered through simple random sampling (SRS).

In a ground-breaking paper, Fay and Herriot (1979) suggested modeling a transform of the weighted estimator via a linear mixed model.

Fay-Herriot Model

Specifically, let y_i denote a transformation of the weighted estimator:

$$y_i = g\left(\underbrace{\frac{\sum_{k \in S_i} w_k y_k}{\sum_{k \in S_i} w_k}}_{\text{Weighted Estimate}}\right).$$

and assume

$$y_i | \beta_0, \delta_i \sim \text{N}(\beta_0 + \boldsymbol{x}_i^\top \boldsymbol{\beta}_1 + \delta_i, V_i),$$

with V_i assumed known and being derived from the variance of the weighted estimator, and

$$\delta_i | \sigma_\delta^2 \sim_{iid} \text{N}(0, \sigma_\delta^2).$$

Rather than depending completely on the covariate model (synthetic estimation), we allow an area-specific deviation.

Fay-Herriot Modeling

- The F-H approach is an extremely popular way of producing area-level estimates.

- This formulation avoids the need to model the design.

- For a continuous outcome (e.g., child growth measures) it may suffice to simply take the weighted estimator, with its associated variance.

- For prevalences, we can work a transform of the weighted prevalence estimates, $\widehat{p}_i$:

$$y_i = g(\widehat{p}_i) = \log\left(\frac{\widehat{p}_i}{1-\widehat{p}_i}\right),$$

  and then use the delta method to obtain the appropriate variance.

Fay-Herriot Modeling

- For rates, we can work with a transform of the weighted rate estimates $\widehat{r}_i$:
$$y_i = g(\widehat{r}_i) = \log(\widehat{r}_i + c),$$
where the constant c is chosen to avoid taking the log of zero, and again use the delta method to obtain the appropriate variance.

- Area-level covariates are frequently used – one way of viewing a F-H model is to start with a synthetic estimator but then add on an area-specific random effect to account for bias (so like a composite estimator).

Fay-Herriot Modeling

- A practical difficulties with this approach is that the direct estimates may be on the boundary for a summary parameter that is not on the whole real line.
- For example, in the binary case we may have $\widehat{p}_i$ equal to 0 or 1.
- In this case, the logit will be undefined.
- Further, a transform of the weighted estimator may not share the same design-based properties as the untransformed estimator, such as being design unbiased.
- These problems may be alleviated by using an unmatched sampling and linking model (You and Rao, 2002).
- A second difficulty is that reliable variance estimates V_i may be unavailable, particularly for areas with few samples. In this case, variance smoothing models can be used (Rao and Molina, 2015, Section 6.4.1).

Fay-Herriot Modeling

- We describe the Fay and Herriot (1979) model in the context of estimating a prevalence.
- Let $\widehat{p}_i$ be the weighted estimator of a prevalence p_i, then consider

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

which is on the whole of the real line.

Fay-Herriot Modeling

- The "data" is taken to be y_i and the sampling model is taken as the asymptotic distribution:

$$y_i \sim \mathrm{N}(\theta_i, V_i),$$

where V_i, the variance of the estimator, is known and θ_i is the logit of the prevalence.

- For a non-spatial model, the random effects model is

$$\theta_i = \beta_0 + \boldsymbol{x}_i^{\mathsf{T}}\boldsymbol{\beta}_1 + \delta_i,$$

where $\boldsymbol{x}_i$ are area-level covariates and the random effects have distribution $\delta_i \sim_{iid} \mathrm{N}(0, \sigma_\delta^2)$.

- The model acknowledges the design and also smooths to a global level – it is straightforward to add spatial random effects.

Spatial Fay-Herriot Modeling

- We now extend to a spatial version, the so-called BYM2 model (Besag *et al.*, 1991; Riebler *et al.*, 2016).

- The model is

$$\theta_i = \beta_0 + \boldsymbol{x}_i^{\top}\boldsymbol{\beta}_1 + \delta_i$$

  with the random effects defined to smooth over space:

$$\delta_i = \sigma_\delta[\sqrt{1-\phi}e_i + \sqrt{\phi}S_i]$$

  where e_i are IID and S_i are spatial random effects[3].

- This is a spatial area-level Fay-Herriot SAE model.

[3] specifically, in SUMMER these are taken intrinsic conditional autoregressive (ICAR)

Simulated Normal Example

Motivating Example: Normal Data

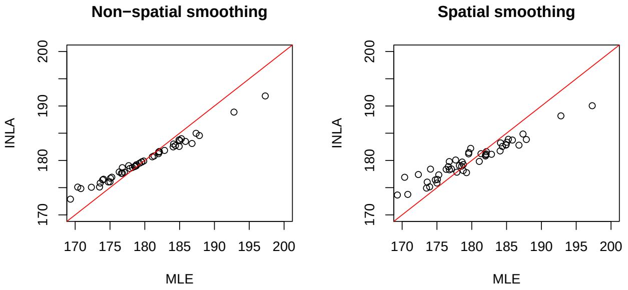

Figure 5: Comparison of area averages: Posterior medians from non-spatial model (described in Lecture 3) versus MLEs (left). Posterior medians from spatial model versus MLEs (right).

The shrinkage is less predictable with the spatial model, which is because of the local adaptation.

Motivating Example: Normal Data

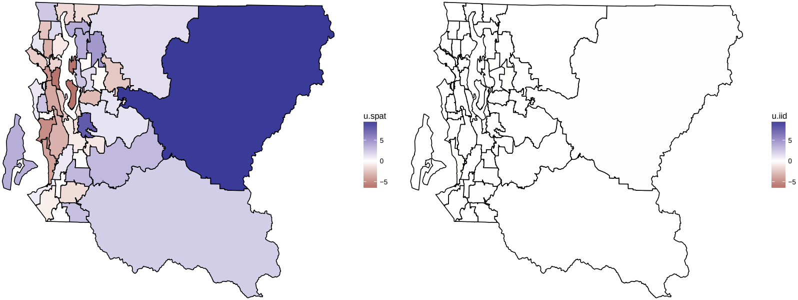

Figure 6: Spatial (left) and non-spatial (left) random effects from the spatial+IID model.

The IID contribution is much smaller than the spatial contribution.

Motivating Example: Normal Data

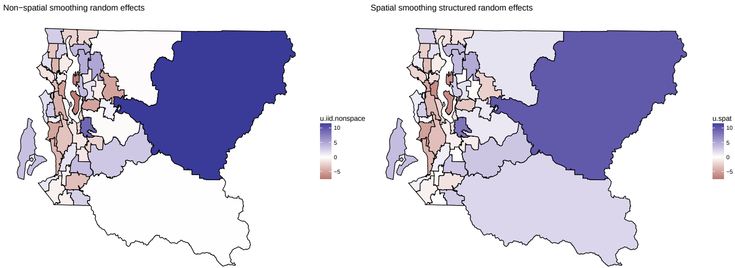

Figure 7: Non-spatial random effects δ_i from the non-spatial model (left) and spatial random effects (right) random effects S_i.

The non-spatial model random effects are trying to pick up the spatial structure!

Motivating Example: Normal Data

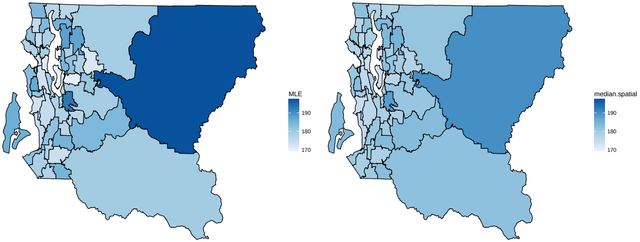

Figure 8: Estimates of area averages of weight via MLE's (left) and posterior medians from spatial model (right).

The extremes are attenuated under the spatial model.

Motivating Example: Normal Data

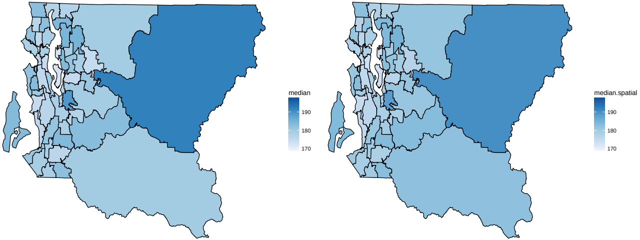

Figure 9: Posterior median estimates of area averages of weight via non-spatial hierarchical model with $\beta_0 + \delta_i$ (left) and spatial hierarchical model $\beta_0 + \delta_i + S_i$ (right); δ_i are iid and S_i are spatial random effects.

Some differences between the estimates, but relatively minor.

The SAR Spatial Model

The SAR Model

Following Marhuenda *et al.* (2013), in the `sae` package, a simultaneous autoregressive (SAR) Fay-Herriot model is available, as an alternative to the BYM2 model (which is in `SUMMER`).

The SAR model is given by,

$$y_i = \boldsymbol{x}_i \boldsymbol{\beta} + \sum_{j=1}^{n} B_{ij}(y_j - \boldsymbol{x}_j \boldsymbol{\beta}) + \epsilon_i,$$

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

- the SAR coefficients B_{ij} are such that $B_{ii} = 0$,
- ϵ_i are independent, zero mean errors terms with variance V_i,
- Let $\text{var}(\boldsymbol{\epsilon}) = \boldsymbol{V}$ be the $n \times n$ diagonal matrix with i-th element V_i.

SAR Models

In vector form:

$$\boldsymbol{y} - \boldsymbol{x}\beta = \boldsymbol{B}(\boldsymbol{y} - \boldsymbol{x}\beta) + \epsilon,$$

or

$$(\mathrm{I}_n - \boldsymbol{B})(\boldsymbol{y} - \boldsymbol{x}\beta) = \epsilon,$$

or

$$\boldsymbol{y} = \boldsymbol{x}\beta + (\mathrm{I}_n - \boldsymbol{B})^{-1}\epsilon,$$

where the latter assumes that $\boldsymbol{B}$ has been chosen so that $\mathrm{I}_n - \boldsymbol{B}$ is invertible.

If $\boldsymbol{B} = \mathbf{0}$ we have an iid model.

So spatial dependence is induced through $(\mathrm{I}_n - \boldsymbol{B})^{-1}$.

SAR Models

Under normality of the errors

$$\boldsymbol{Y} \sim \mathrm{N}(\boldsymbol{x}\boldsymbol{\beta}, [(\mathrm{I}_n - \boldsymbol{B})^\top \boldsymbol{V}(\mathrm{I}_n - \boldsymbol{B})^{-1}]^\top).$$

For a well-defined model we require $(\mathrm{I}_n - \boldsymbol{B})$ to be non-singular, which puts conditions on $\boldsymbol{B}$.

A common choice is

$$\boldsymbol{B} = \rho_s \boldsymbol{W},$$

for a spatial proximity matrix $\boldsymbol{W}$ with elements 1 for neighbors and 0 otherwise.

SAR Models: Some Technical Details

Let $\lambda_{(1)}, \ldots, \lambda_{(n)}$ be the ordered eigenvalues of $\boldsymbol{W}$.

Then, $(\mathrm{I}_n - \boldsymbol{B})$ will be invertible if

$$\frac{1}{\lambda_{(1)}} < \rho_s < \frac{1}{\lambda_{(n)}}.$$

If the row sums of $\boldsymbol{W}$ are standardized to 1 then $\lambda_{(n)} = 1$ and $\lambda_{(1)} \leq -1$ so $\rho_s < 1$ but may be less than -1.

For more discussion, see Banerjee et al. (2015, Section 4.4) and Schabenberger and Gotway (2005, Section 6.2.2.1).

Fay-Herriot with a SAR Model

- 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 SAR spatial model (Marhuenda *et al.*, 2013) was used.

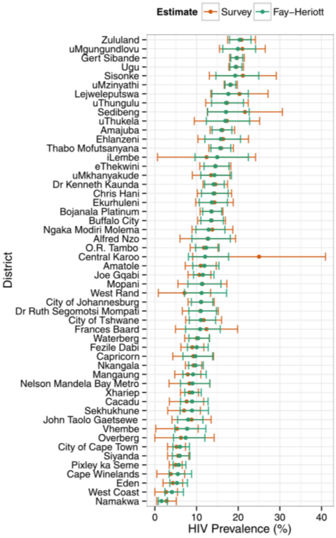

Figure 10: Direct and Fay-Herriot estimates of HIV prevalence in South African districts in 2012, from Gutreuter *et al.* (2019).

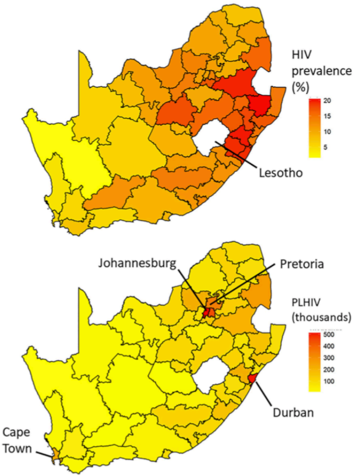

Figure 11: Estimates of HIV prevalence and people living with HIV in South African districts in 2012, from Gutreuter *et al.* (2019).

Spatio-Temporal Smoothing

Main Effects and Interactions

- To motivate space-time models, when space is modeled discretely, we consider simple two-way factor models.
- Suppose we have a univariate continuous response Y.
- Suppose we have two factors, A and B say, with $i = 1, \ldots, m$ and $j = 1, \ldots, T$ indexing the levels.
- A main effects only model takes the form

$$E[Y_{ij}|\beta_0, \eta_i, \delta_j] = \beta_0 + \eta_i + \tau_j.$$

- **Interpretation:** η_i is the effect of being at level i for factor A, regardless of the level assumed by B, and τ_j is the effect of being at level j for factor B, regardless of the level assumed by A, i.e. there is no interaction.

Main Effects and Interactions

- An interaction model adds a set of interaction parameters

$$E[Y_{ij}|\beta_0, \eta_i, \tau_j, \delta_{ij}] = \beta_0 + \eta_i + \tau_j + \delta_{ij}.$$

- **Interpretation:** δ_{ij} is the additional effect, beyond $\eta_i + \tau_j$ of being simultaneously at levels *i* and *j* of factors A and B.

- If the factor correspond to nominal levels (e.g., a factor for color with 2 levels: ”red”, ”blue”) then we would not expect similarity between adjacent levels.

- In a space-time context the “factors” space and time have an “ordering” and we might expect similarity.

Main Effects Model

- First, consider the space-time model for a binary outcome,

$$\begin{aligned} Y_{it}|\theta_{it} &\sim \text{N}(\theta_{it}, V_{it}) \\ \theta_{it} &= \text{expit}(\beta_0 + \eta_i + S_i + \omega_t + \tau_t) \end{aligned}$$

- Components:
  - Y_{it} is the weighted estimate with associated design-based variance V_{it}.
  - Unstructured spatial term $\eta_i \sim_{iid} \text{N}(0, \sigma_e^2), i = 1, \ldots, m$.
  - Smooth spatial term $[S_1, \ldots, S_m]$ smooth in space, e.g., from an ICAR model.
  - Unstructured temporal term $\omega_t \sim_{iid} \text{N}(0, \sigma_\omega^2), t = 1, \ldots, T$.
  - Smooth temporal term $[\tau_1, \ldots, \tau_T]$ smooth in time, e.g. follows an RW1 or RW2 model.
- Notice there is no interaction between space and time.
- The spatial effects are constant across time and temporal trends are constant across space.

Space-Time Interaction Models

- Knorr-Held (2000) considered the model:

$$\theta_{it} = \beta_0 + \eta_i + S_i + \omega_t + \tau_t + \delta_{it},$$

  with η_i, S_i, ω_t, δ_t are as in the main effects only model.

- Four different models for the interaction δ_{it}:

  - Type I: Independent interaction.
  - Type II: Temporal trends differ between areas but don't have spatial structure.
  - Type III: Spatial patterns differ between time points but don't have temporal structure.
  - Type IV: Temporal trends differ between areas but more likely to be similar for adjacent areas.

We describe the Type IV model only, since it is the most appealing in a prevalence mapping context.

Inseparable Space-Time Interaction Models

- Type IV: Temporal trends differ between areas but more likely to be similar for neighboring areas.

- This will often be the most realistic model if interactions are present.

- In the case of a RW2 temporal model and an ICAR spatial model, the joint distribution can be written:

$$p(\boldsymbol{\delta}|\sigma_\delta^2) \propto \exp\left(-\frac{1}{2\sigma_\delta^2}\sum_{t=3}^{T}\sum_{i\sim j}(\delta_{it}-\delta_{jt}-2\delta_{i,t-1}+2\delta_{j,t-1}+\delta_{i,t-2}-\delta_{j,t-2})^2\right)$$

- The Knorr-Held (2000) models are implemented in the `SUMMER` package.

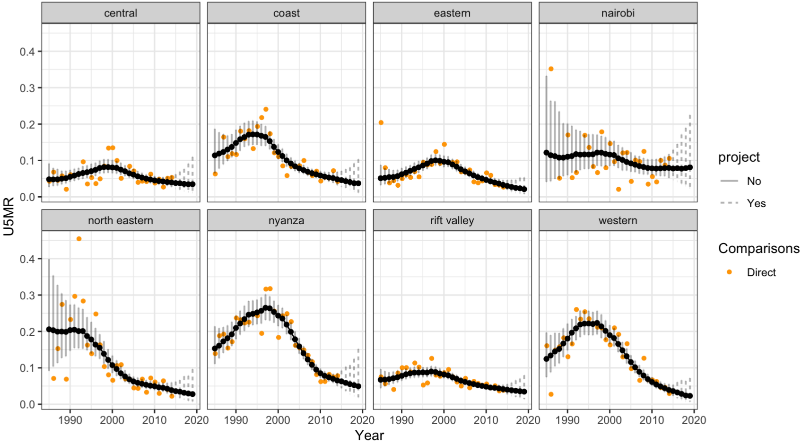

Figure 12: Weighted estimates and smoothed fits over time for 8 provinces of Kenya.

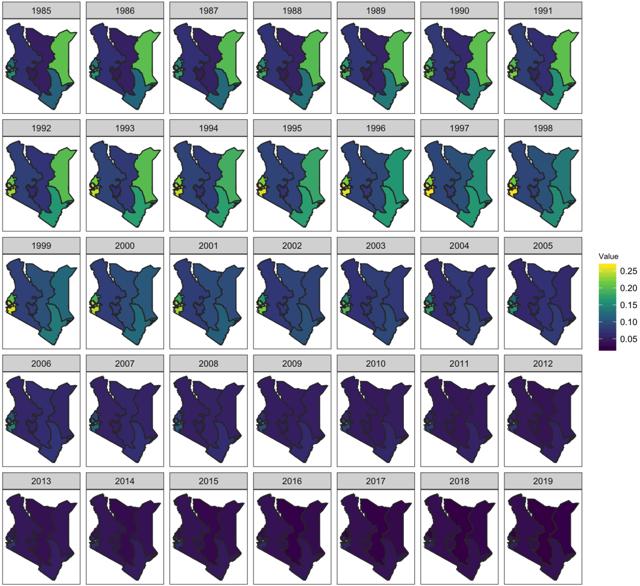

Figure 13: Maps of smoothed estimates over time for 8 provinces of Kenya.

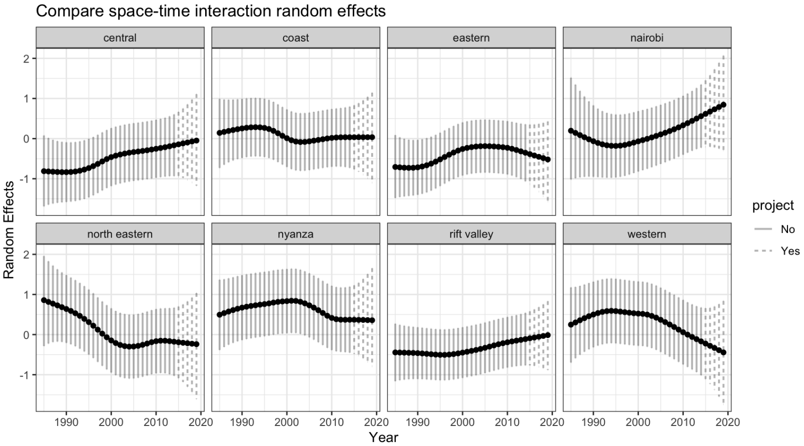

Figure 14: Space-time interactions δ_{it} for 8 provinces of Kenya.

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

- The space-time Fay-Herriot 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.
- This smoothed direct model is very reliable for examining Admin1 subnational variation, but the direct estimates are often unreliable for Admin2 estimation – hence, in the next lecture, we describe a unit-level model for this endeavor.

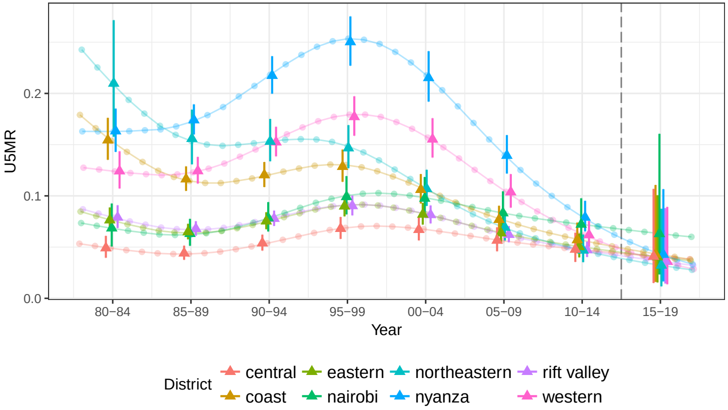

Figure 15: Posterior median estimates for Kenya districts.

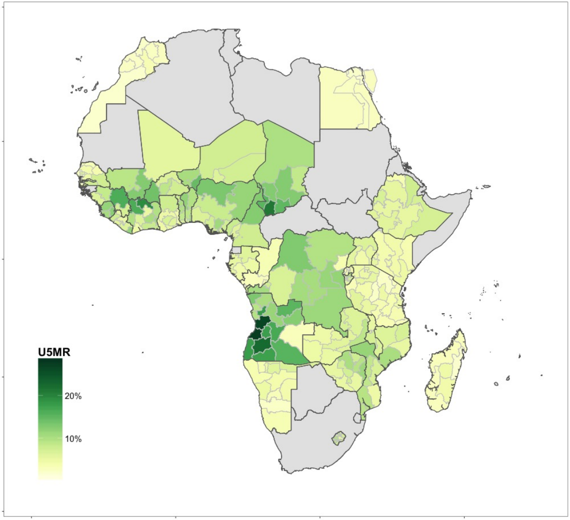

Figure 16: Predictions of U5MR for 2015, in 35 countries of Africa.

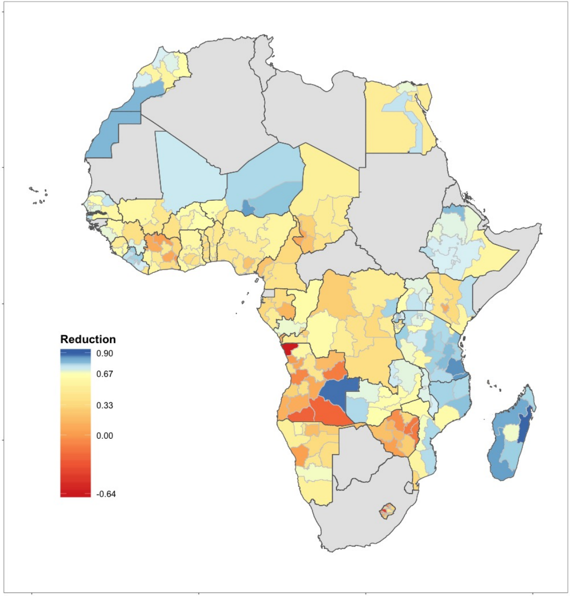

Figure 17: Percent reduction from 1990 to 2015, in 35 countries of Africa.

Discussion

Discussion

- If the data are sparse in an area, averages and totals are unstable because of the small denominators.
- More reliable estimates can be obtained by using the totality of data to inform on the distribution, both locally and globally, of the averages across the study region.
- A LMEM can include spatial dependence relatively easily, with the ICAR model being particularly popular.

Discussion

Four levels of understanding for hierarchical models, in descending order of importance:

- The intuition on global and local smoothing.
- The models to achieve this.
- How to specify prior distributions.
- The computation behind the modeling.

Overall Strategy

- First, calculate empirical means and map them. Also look at map of standard errors and/or confidence intervals.
- Fit non-spatial random effects models.
- Fit the ICAR+IID spatial model.
- Add in covariates if available.

Discussion: Area-Level Modeling

- Mixed effects models are a common approach for SAE, as they acknowledge between-area differences in outcomes.
- But one must consider the design (e.g., stratification and clustering) – Fay-Herriot modeling is a simple way to do this.
- Model-checking is important.

References

Banerjee, S., Carlin, B., and Gelfand, A. (2015). *Hierarchical Modeling and Analysis for Spatial Data, Second Edition*. CRC Press.

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

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

Burris, K. and Hoff, P. (2020). Exact adaptive confidence intervals for small areas. *Journal of Survey Statistics and Methodology*, **8**, 206–230.

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.

Krainski, E. T., Gómez-Rubio, V., Bakka, H., Lenzi, A., Castro-Camilo, D., Simpson, D., Lindgren, F., and Rue, H. (2018). *Advanced Spatial Modeling with Stochastic Partial Differential Equations Using R and INLA*. Chapman and Hall/CRC.

Li, Z. R., Hsiao, Y., Godwin, J., Martin, B. D., Wakefield, J., and Clark, S. J. (2019). Changes in the spatial distribution of the under five mortality rate: small-area analysis of 122 DHS surveys in 262 subregions of 35 countries in Africa. *PLoS One*, **14**, e0210645.

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

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.

Otto, M. C. and Bell, W. R. (1995). Sampling error modelling of poverty and income statistics for states. In *American Statistical Association, Proceedings of the Section on Government Statistics*, pages 160–165.

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

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

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

Schabenberger, O. and Gotway, C. A. (2005). *Statistical Methods for Spatial Data Analysi*. CRC press.

Wang, X., Yue, Y., and Faraway, J. J. (2018). *Bayesian Regression Modeling with INLA*. Chapman and Hall/CRC.

You, Y. and Rao, J. (2002). Small area estimation using unmatched sampling and linking models. *Canadian Journal of Statistics*, **30**, 3–15.