Session 7.1: Spatial models for small area data: disease mapping and ecological regression

Bayesian modelling for Spatial and Spatio-temporal data, Imperial College

Learning Objectives

After this session you should be able to:

- Explain the main ideas underlying the use of Bayesian methods for producing spatially smoothed estimates of disease risk in small areas
- Describe different priors for spatial random effects
- Explore aetiological hypothesis between a health outcome and exposure based on disease mapping
- Describe Poisson regression with spatial random effects for continuous and categorical covariates;
- Use `R-INLA` to produce maps of smoothed estimates of disease risk, carry out spatial smoothing of disease risk and specify ecological regression models

The topics treated in this lecture are covered in Chapter 5-6 of Blangiardo and Cameletti (2015)

Outline

1. Spatial Structure

2. Example: suicides in London

3. Ecological regression with spatial random effects

Smoothed estimates of the RR (non spatial)

- Poisson-logNormal model based on the assumption that the observations in the data set are identically distributed and independent
- However, data that occur close together in space (or time) are likely to be correlated
  → Dependence between observations is a more realistic assumption
- Ignoring this dependence can lead to biased and inefficient inference
  → Smooth in space prior distribution for the random effects should allow for spatial correlation

Spatial structure

Intrinsic CAR model (Besag, York, and Mollie, 1991)

- Specify the distribution of each random effect as if we knew the values of the spatial random effects in neighbouring areas
- We have a conditional specification since we are conditioning on knowing the neighbours
- Rule for determining the neighbours of each area: most common based on common boundary
- Use of conditional autoregressive distributions

General definition | Common definition | Remarks

$$\mathbf{u} \sim \text{ICAR}(\mathbf{W}, \sigma_u^2)$$

with

- $\mathbf{W}$ matrix defining the neighbours (weights)
- σ_u^2 conditional variance parameter of $\mathbf{U}$

$$u_i \mid u_{j \ \ j \neq i} \sim \text{Normal}\left(\frac{\sum_j W_{ij} u_j}{\sum_j W_{ij}}, \frac{\sigma_u^2}{\sum_j W_{ij}}\right)$$

Intrinsic CAR model (Besag, York, and Mollie, 1991)

- Specify the distribution of each random effect as if we knew the values of the spatial random effects in neighbouring areas
- We have a conditional specification since we are conditioning on knowing the neighbours
- Rule for determining the neighbours of each area: most common based on common boundary
- Use of conditional autoregressive distributions

General definition | **Common definition** | Remarks

$$\mathbf{u} \sim \text{ICAR}(\mathbf{W}, \sigma_u^2)$$

Let $\partial_i =$ set of areas adjacent to i, w_{ij} = 1 for $j \in \partial_i$, 0 otherwise

$$u_i \mid u_{j \neq i} \sim \text{Normal}\left(\frac{\sum_{j \in \partial_i} u_j}{n_i}, \frac{\sigma_u^2}{n_i}\right)$$

- u_i is smoothed towards mean risk in a set of neighbouring areas
- Conditional variance inversely proportional to the number of neighbours (so more neighbours, less variability)

Intrinsic CAR model (Besag, York, and Mollie, 1991)

- Specify the distribution of each random effect as if we knew the values of the spatial random effects in neighbouring areas
- We have a conditional specification since we are conditioning on knowing the neighbours
- Rule for determining the neighbours of each area: most common based on common boundary
- Use of conditional autoregressive distributions

- ICAR model is improper: the overall mean of the $\boldsymbol{u}$ is not defined. So an additional constraints needs to be imposed: **sum-to-zero constraint:** $\sum_i u_i = 0$
- The parameter σ_u^2 represents the **conditional** variance of the random effects (and not the marginal one) and its magnitude determines the amount of spatial variation
- No closed-form expression available for the **marginal** between-area variance of the spatial effects → estimate marginal spatial variance empirically

$$s^2_{\text{u.marginal}} = \sum_i (u_i - \bar{u})^2/(N-1)$$

Combining ICAR with unstructured random effects

- ICAR model makes a strong spatial assumption; it cannot take a limiting form that allows non-spatial variability
- Besag, York and Mollie (BYM) recommended combining the ICAR prior and the standard normal prior to allow for both
  - spatially unstructured latent covariates $\boldsymbol{v}$ modelled as iid → global smoothing
  - spatially correlated latent covariates $\boldsymbol{u}$ modelled as ICAR → local smoothing

BYM | BYM2

$$
\begin{aligned}
y_i &\sim \text{Poisson}(\lambda_i = \rho_i E_i) \\
\eta_i &= \log \rho_i = b_0 + v_i + u_i \\
v_i &\sim \text{Normal}(0, \sigma_v^2) \\
\mathbf{u} &\sim \text{ICAR}(\mathbf{W}, \sigma_u^2)
\end{aligned}
$$

- Need to specify hyperprior distributions for:
- σ_v^2 (between-area unstructured marginal variance), e.g. $1/\sigma_v^2 \sim \text{Gamma}(1, 0.001)$
- σ_u^2 (between-area spatial conditional variance), e.g. $1/\sigma_u^2 \sim \text{Gamma}(1, 0.001)$
- b_0 (mean log relative risk), e.g. $b_0 \sim \text{Normal}(0, 0.0001)$

Combining ICAR with unstructured random effects

- ICAR model makes a strong spatial assumption; it cannot take a limiting form that allows non-spatial variability
- Besag, York and Mollie (BYM) recommended combining the ICAR prior and the standard normal prior to allow for both
  - spatially unstructured latent covariates $\boldsymbol{v}$ modelled as iid → global smoothing
  - spatially correlated latent covariates $\boldsymbol{u}$ modelled as ICAR → local smoothing

BYM | BYM2

$$y_i \sim \text{Poisson}(\lambda_i = \rho_i E_i)$$
$$\eta_i = \log \rho_i = b_0 + b_i$$
$$\boldsymbol{b} = \frac{1}{\sqrt{\tau_b}}(\sqrt{1-\phi}\boldsymbol{v}_* + \sqrt{\phi}\boldsymbol{u}_*)$$

where $\boldsymbol{v}_*$ and $\boldsymbol{u}_*$ are standardised versions of $\boldsymbol{u}$ and $\boldsymbol{v}$.

- Need to specify hyperprior distributions for:
- ϕ which is the weight of the spatially structured residual
- τ_b which is the marginal variance of the random effect

Priors for BYM2

Under the BYM2 specification the hyperparameters τ_b and ϕ are modelled using **Penalised Complexity** (PC) priors (Simpson, Rue, Riebler, Martins, and Sorbye, 2017)

- Regularise inference while not forcing too strong information
- Penalise departure from a "base" model (eg parameter = some fixed value)
- Prior tends to favour the base model $\rightarrow$ need fairly strong evidence to move away from it
- Distance between the **base** model $g(\xi)$ and an **alternative**, more complex model $f(\xi)$ is measured by

$$d(f,g) = \sqrt{2\text{kld}(f,g)} \qquad \text{with} \qquad \text{kld}(f,g) = \int f(\xi)\log\left(\frac{f(\xi)}{g(\xi)}\right) d\xi$$

Priors for BYM2

Under the BYM2 specification the hyperparameters τ_b and ϕ are modelled using **Penalised Complexity** (PC) priors (Simpson, Rue, Riebler, Martins, and Sorbye, 2017)

- Regularise inference while not forcing too strong information
- Penalise departure from a "base" model (eg parameter = some fixed value)
- Prior tends to favour the base model $\rightarrow$ need fairly strong evidence to move away from it
- Distance between the **base** model $g(\xi)$ and an **alternative**, more complex model $f(\xi)$ is measured by

$$d(f,g) = \sqrt{2\text{kld}(f,g)} \qquad \text{with} \qquad \text{kld}(f,g) = \int f(\xi)\log\left(\frac{f(\xi)}{g(\xi)}\right) d\xi$$

- Penalisation done at a constant rate

$$p(d) = \lambda\exp(-\lambda d) \sim \text{Exponential}(\lambda) \qquad \Rightarrow \qquad p(\xi) = \lambda e^{-\lambda d(\xi)} \left|\frac{\partial d(\xi)}{\partial \xi}\right|$$

- PC prior defined using probability statements on the model parameters (in the appropriate scale) to determine the value of λ using "reasonable" information

Priors for BYM2

Using probability statements we can define the PC priors for the two hyperparameters as:

Prior on τ_b

$$P((1/\sqrt{\tau_b}) > U_1) = \alpha_1$$

which can be interpreted as *the probability that the standard deviation of the random effect is larger than U_1 is equal to α_1*

Prior on ϕ

$$P(\phi < U_2) = \alpha_2$$

which can be interpreted as *the probability that the spatial random effect explains less than U_2 of the total variability is equal to α_2*

Priors for BYM2

Using probability statements we can define the PC priors for the two hyperparameters as:

Prior on τ_b

$$P((1/\sqrt{\tau_b}) > U_1) = \alpha_1$$

which can be interpreted as *the probability that the standard deviation of the random effect is larger than U_1 is equal to α_1*

Prior on ϕ

$$P(\phi < U_2) = \alpha_2$$

which can be interpreted as *the probability that the spatial random effect explains less than U_2 of the total variability is equal to α_2*

where $\lambda = \frac{-\log(\alpha_k)}{U_k}$

- We need to define U_1, U_2 and α_1, α_2 and following Simpson, Rue, Riebler, et al. (2017):

Priors for BYM2

Using probability statements we can define the PC priors for the two hyperparameters as:

Prior on τ_b

$$P((1/\sqrt{\tau_b}) > U_1) = \alpha_1$$

which can be interpreted as *the probability that the standard deviation of the random effect is larger than U_1 is equal to α_1*

Prior on ϕ

$$P(\phi < U_2) = \alpha_2$$

which can be interpreted as *the probability that the spatial random effect explains less than U_2 of the total variability is equal to α_2*

where $\lambda = \frac{-\log(\alpha_k)}{U_k}$

- We need to define U_1, U_2 and α_1, α_2 and following Simpson, Rue, Riebler, et al. (2017):
  - A marginal sd not too large (e.g. 0.5);

Priors for BYM2

Using probability statements we can define the PC priors for the two hyperparameters as:

Prior on τ_b

$$P((1/\sqrt{\tau_b}) > U_1) = \alpha_1$$

which can be interpreted as *the probability that the standard deviation of the random effect is larger than U_1 is equal to α_1*

Prior on ϕ

$$P(\phi < U_2) = \alpha_2$$

which can be interpreted as *the probability that the spatial random effect explains less than U_2 of the total variability is equal to α_2*

where $\lambda = \frac{-\log(\alpha_k)}{U_k}$

- We need to define U_1, U_2 and α_1, α_2 and following Simpson, Rue, Riebler, et al. (2017):
  - A marginal sd not too large (e.g. 0.5);
  - $\alpha_1 = 0.01$ (we want to allow for a small probability)

Priors for BYM2

Using probability statements we can define the PC priors for the two hyperparameters as:

Prior on τ_b

$$P((1/\sqrt{\tau_b}) > U_1) = \alpha_1$$

which can be interpreted as *the probability that the standard deviation of the random effect is larger than U_1 is equal to α_1*

Prior on ϕ

$$P(\phi < U_2) = \alpha_2$$

which can be interpreted as *the probability that the spatial random effect explains less than U_2 of the total variability is equal to α_2*

where $\lambda = \frac{-\log(\alpha_k)}{U_k}$

- We need to define U_1, U_2 and α_1, α_2 and following Simpson, Rue, Riebler, et al. (2017):
  - A marginal sd not too large (e.g. 0.5);
  - $\alpha_1 = 0.01$ (we want to allow for a small probability)
  - $\alpha_2 = 2/3$ (we expect a higher probability that the variability to be explained by the spatial random effect is lower than 50%)

Priors for BYM2

Using probability statements we can define the PC priors for the two hyperparameters as:

Prior on τ_b

$$P((1/\sqrt{\tau_b}) > U_1) = \alpha_1$$

which can be interpreted as *the probability that the standard deviation of the random effect is larger than U_1 is equal to α_1*

Prior on ϕ

$$P(\phi < U_2) = \alpha_2$$

which can be interpreted as *the probability that the spatial random effect explains less than U_2 of the total variability is equal to α_2*

where $\lambda = \frac{-\log(\alpha_k)}{U_k}$

- We need to define U_1, U_2 and α_1, α_2 and following Simpson, Rue, Riebler, et al. (2017):
  - A marginal sd not too large (e.g. 0.5);
  - $\alpha_1 = 0.01$ (we want to allow for a small probability)
  - $\alpha_2 = 2/3$ (we expect a higher probability that the variability to be explained by the spatial random effect is lower than 50%)
- Then
  1. Assuming $U_1 = 0.5/0.31$ translates into $P(\sigma_{\tau_b} > 1.62) = 0.01$, using the rule of thumb in Simpson, Rue, Riebler, et al. (2017)
  2. Assuming $U_2 = 0.5$ we get $P(\phi < 0.5) = 2/3$

Poisson model with BYM random effects

- Choice of the adjacency matrix (neighbours): 2 areas are neighbours if they share a common border
  → Adjacency matrix implemented in INLA
- An area cannot be specified as its own neighbour
- Adjacency matrix must be symmetric

Poisson model with BYM random effects

- Choice of the adjacency matrix (neighbours): 2 areas are neighbours if they share a common border
  $\rightarrow$ Adjacency matrix implemented in INLA
- An area cannot be specified as its own neighbour
- Adjacency matrix must be symmetric
- $\mathrm{RR}_i = \exp(b_0 + b_i)$: RR in area i relative to the age/sex structure (used to estimate the E_i)
- $\text{residual } \mathrm{RR}_i = \exp(b_i)$: residual RR in area i relative to the region average after adjusting for the overall risk

Poisson model with BYM random effects

- Choice of the adjacency matrix (neighbours): 2 areas are neighbours if they share a common border
  $\rightarrow$ Adjacency matrix implemented in INLA
- An area cannot be specified as its own neighbour
- Adjacency matrix must be symmetric
- $\mathrm{RR}_i = \exp(b_0 + b_i)$: RR in area i relative to the age/sex structure (used to estimate the E_i)
- $\text{residual } \mathrm{RR}_i = \exp(b_i)$: residual RR in area i relative to the region average after adjusting for the overall risk
- σ_b^2 reflects the marginal variability of the REs
- ϕ represent the weight of the spatial structure

Example: Suicides in London

Suicides in Greater London, M+F, 1989-1993, Boroughs

- 32 boroughs in Greater London
- Interest: mapping the RR in each borough
- Methods with no spatial structure: SMR, non spatial smoothing
- Spatial smoothing using the BYM model

$$
\begin{aligned}
y_i &\sim \text{Poisson}(\rho_i E_i) \\
\log \rho_i &= b_0 + b_i \\
v_i &\sim \text{Normal}(0, \sigma_v^2) \\
\boldsymbol{u} &\sim \text{ICAR}(\mathbf{W}, \sigma_u^2)
\end{aligned}
$$

- Data: $\boldsymbol{y}$ and $\boldsymbol{E}$
- Priors: σ_v^2, σ_u^2, b_0
- Parameters of interest:
  – residual RR $(\text{resRR}_i = \exp(b_i))$
  – marginal variance $(1/\tau_b)$
  – percent of total variation in the log RR due to spatial effects (ϕ)

Adjacency matrix in INLA

- It is possible to produces a graph from a shapefile
- Upload the shapefile using `sf` package

```
> library(sf)
> london.gen <- read_sf("LDNSuicides.shp")
> london.gen$ID <- seq(1,32)
```

Adjacency matrix in INLA

- It is possible to produces a graph from a shapefile
- Upload the shapefile using `sf` package

```
> library(sf)
> london.gen <- read_sf("LDNSuicides.shp")
> london.gen$ID <- seq(1,32)
```

- Use `poly2nb` and `nb2INLA` from the `spdep` package to transform the shapefile into adjacency matrix

```
> library(spdep)
> nb2INLA("LDN.graph",poly2nb(london.gen))
> LDN.adj <- paste(getwd(),"/LDN.graph",sep="")
```

Adjacency matrix in INLA

- It is possible to produces a graph from a shapefile
- Upload the shapefile using `sf` package

```
> library(sf)
> london.gen <- read_sf("LDNSuicides.shp")
> london.gen$ID <- seq(1,32)
```

- Use `poly2nb` and `nb2INLA` from the `spdep` package to transform the shapefile into adjacency matrix

```
> library(spdep)
> nb2INLA("LDN.graph",poly2nb(london.gen))
> LDN.adj <- paste(getwd(),"/LDN.graph",sep="")
```

- Now `LDN.graph` has been saved in the working directory and can be called when specifying the BYM model (see later)

Spatial distributions for area level data in INLA

We introduce here the specification of the `ICAR` and `BYM2` models in `INLA`, which are done through `f()`:

ICAR in INLA | BYM2 in INLA

```
> formula.ICAR <- y ~ f(ID, model="besag", graph=LDN.adj)
```

- `ID` is the area identifier
- `graph=LDN.adj` identifies the adjacency structure constructed as seen before
- `model=besag` specifies the intrinsic conditional autoregressive structure as described before

On the example:

```
> formula <- y ~ 1 + f(ID, model="besag",graph=LDN.adj,
+                      hyper=list(
+                        prec=list(
+                          prior="loggamma",param=c(1,0.0005))))
```

Spatial distributions for area level data in INLA

We introduce here the specification of the `ICAR` and `BYM2` models in `INLA`, which are done through `f()`:

ICAR in INLA | **BYM2 in INLA**

```
> formula.BYM2 <- y ~ f(ID, model="bym2", graph=LDN.adj)
```

- `ID` is the area identifier
- `graph=LDN.adj` identifies the adjacency structure constructed as seen before
- `model=BYM2` specifies the combination of the intrinsic conditional autoregressive structure and unstructured random effect as described before

On the example:

```
> formula <- y ~ 1 + f(ID, model="bym2",graph="LDN.graph",
+         hyper=list(prec = list(
+         prior = "pc.prec",
+         param = c(0.5 / 0.31, 0.01)),
+         phi = list(
+         prior = "pc",
+         param = c(0.5, 2 / 3))))
```

Running the model in INLA

To run the model in INLA

```
> mod.suicides <- inla(formula,family="poisson",
+                      data=data.suicides,E=E,
+                      control.compute=list(dic=TRUE, waic=TRUE))
```

R-INLA estimates the parameters $\boldsymbol{\theta} = \{b_0, \boldsymbol{b}, \boldsymbol{u}\}$ and the hyper-parameters $\boldsymbol{\psi} = \{\tau_b, \phi\}$.

How to get information from random effects

- The random effect are obtained through

```
> mod.suicides$summary.random$ID
```

```
  ID        mean         sd  0.025quant    0.5quant  0.975quant        mode
1  1 -0.08520171 0.10818816 -0.30316548 -0.08327256  0.12195201 -0.07943043
2  2 -0.17475921 0.08549566 -0.34620546 -0.17355028 -0.01009475 -0.17112215
3  3 -0.21985243 0.09674533 -0.41433934 -0.21832516 -0.03396347 -0.21530631
4  4  0.12066795 0.08256690 -0.04279373  0.12104943  0.28184663  0.12182984
5  5 -0.14177226 0.08409894 -0.30941934 -0.14097440  0.02131868 -0.13937534
6  6  0.40346274 0.08264884  0.23919012  0.40410431  0.56410646  0.40545398
           kld
1 4.362964e-06
2 2.551789e-06
3 1.615018e-06
4 3.587001e-06
5 2.827558e-06
6 2.128593e-06
```

which is a matrix formed by $2n$ rows:

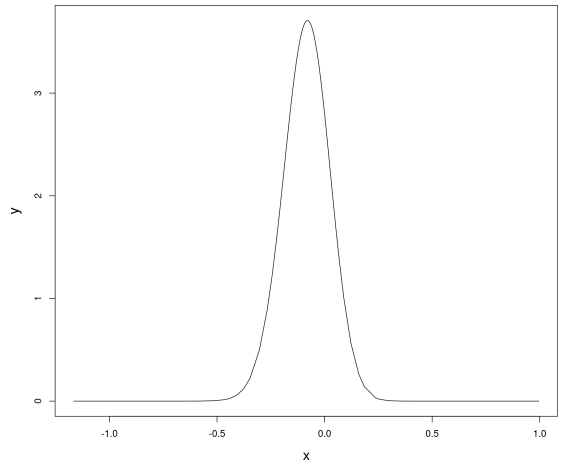
- $1 : n$ rows include information on the area specific residuals b_i
- $n + 1 : 2n$ rows are the spatially structured residual u_i

How to get information from random effects

- All these parameters are on the logarithmic scale; to transform the **marginal** back to the natural scale:

```
> b <- mod.suicides$marginals.random$ID[1:Narea
```

this returns a list with `Nareas` number of elements, each representing the posterior marginal of b_i for that area

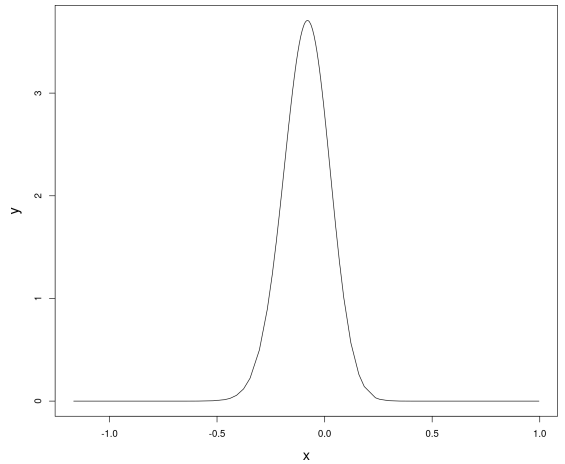


How to get information from random effects

- All these parameters are on the logarithmic scale; to transform the **marginal** back to the natural scale:

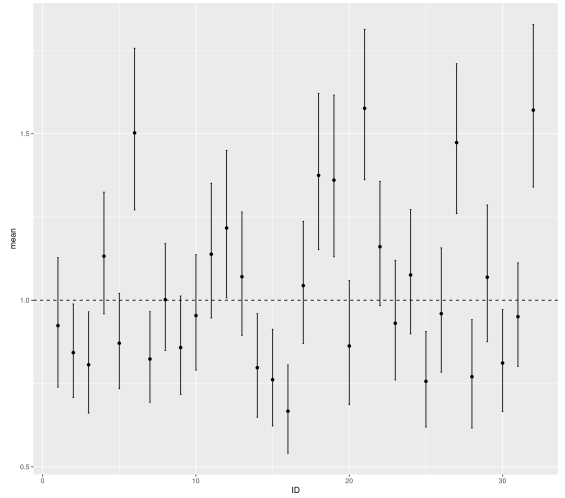
```
> b <- mod.suicides$marginals.random$ID[1:Narea
```

this returns a list with `Nareas` number of elements, each representing the posterior marginal of b_i for that area



- Then we can get the posterior mean and 95% credible intervals:

```
> zeta <- lapply(b,function(x) inla.emarginal(exp,x))
> zeta_CI <- lapply(b,function(x)
+ inla.qmarginal(c(0.025,0.975),
+     inla.tmarginal(exp,x)))
```



Identification of spatial patterns

- What is the sensitivity vs specificity of smoothed RR?
  – Ability to detect true patterns (sensitivity)
  – Ability to discard false patterns (specificity)

- Detection of increased/decreased RR

  → Posterior probabilities that the residual RR is above/below 1 (Richardson, Thomson, Best, and Elliott, 2004)

- Area with an increased risk

$$\begin{aligned} \mathrm{P}(\mathrm{resRR}_i > 1) > 0.8 &\Leftrightarrow \mathrm{P}(e^{(u_i+v_i)} > 1) > 0.8 \\ &\Leftrightarrow \mathrm{P}(u_i + v_i > 0) > 0.8 \end{aligned}$$

- Area with a decreased risk

$$\begin{aligned} \mathrm{P}(\mathrm{resRR}_i < 1) > 0.8 &\Leftrightarrow \mathrm{P}(e^{(u_i+v_i)} > 1) < 0.2 \\ &\Leftrightarrow \mathrm{P}(u_i + v_i > 0) < 0.2 \end{aligned}$$

Posterior probability in INLA

- Remember the parametrisation $\zeta = \exp(b_i)$
- We can visualize $p(\zeta_i > 1 \mid \boldsymbol{y}) = p(b_i > 0 \mid \boldsymbol{y})$ using the built-in function `inla.pmarginal`:

```
> a <- 0
> prob.b <- lapply(b, function(x) {1 - inla.pmarginal(a, x)})
```

Posterior probability in INLA

- Remember the parametrisation $\zeta = \exp(b_i)$
- We can visualize $p(\zeta_i > 1 \mid \boldsymbol{y}) = p(b_i > 0 \mid \boldsymbol{y})$ using the built-in function `inla.pmarginal`:

```
> a <- 0
> prob.b <- lapply(b, function(x) {1 - inla.pmarginal(a, x)})
```

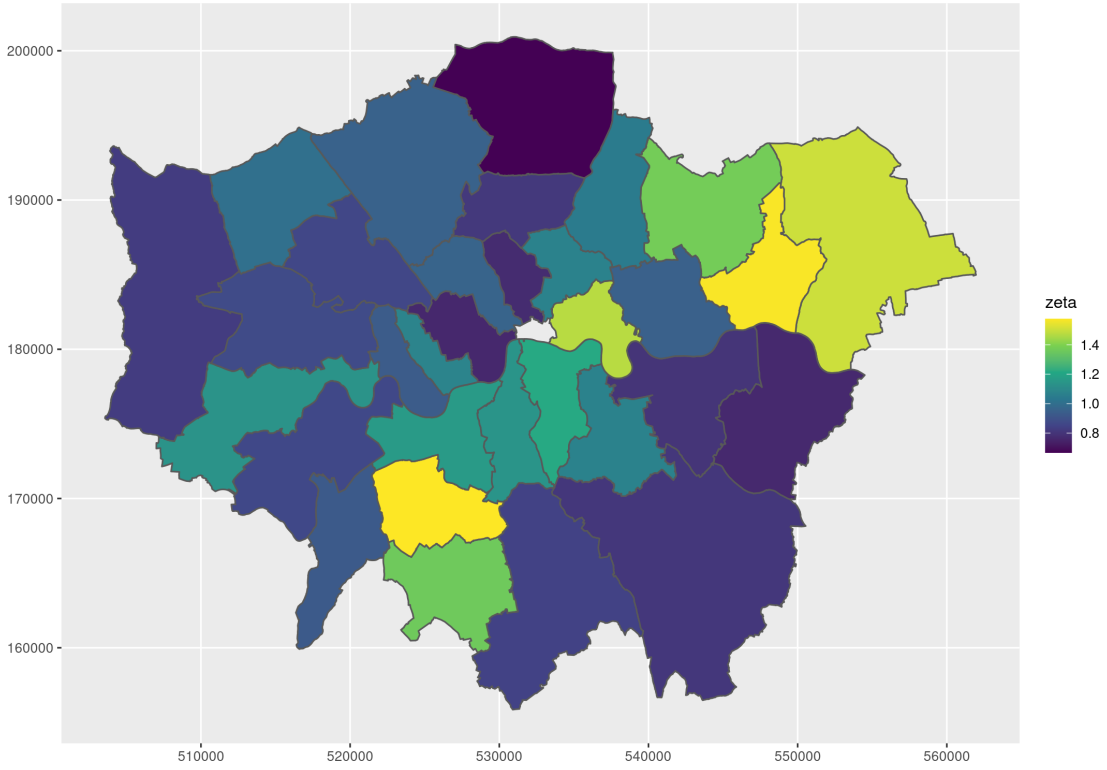
- Create an object with all the info to map

```
> RR_BYM <- tibble(zeta=unlist(zeta),prob=unlist(prob.b), ID=seq(1,32))
> out_map <- left_join(london.gen,datasuicides, by="ID") %>% left_join(., RR_BYM, by="ID")
> out_map$pp_breaks <- cut(out_map$prob,
+                          breaks = c(0, c(0.2, 0.8),
+                                     1), include.lowest = T)
```

...and then create some maps

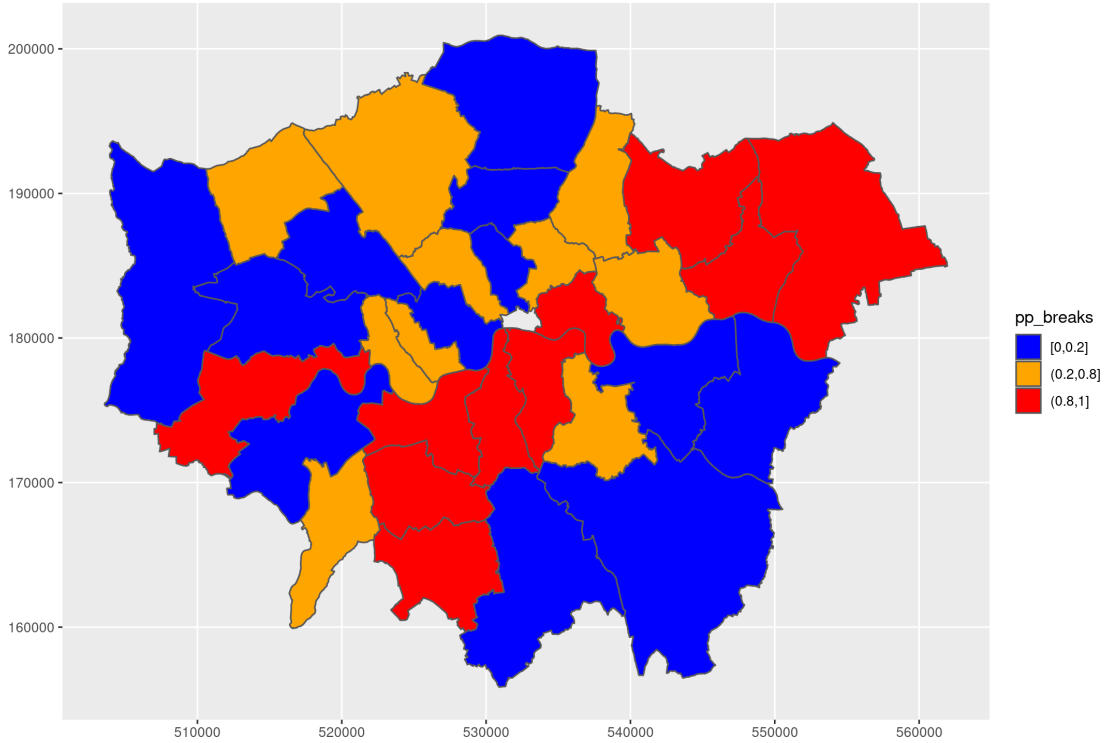
Map of posterior mean of b_i

```
> ggplot() + geom_sf(data = out_map,
+                    aes(fill = zeta)) +
+              scale_fill_viridis_c()
```



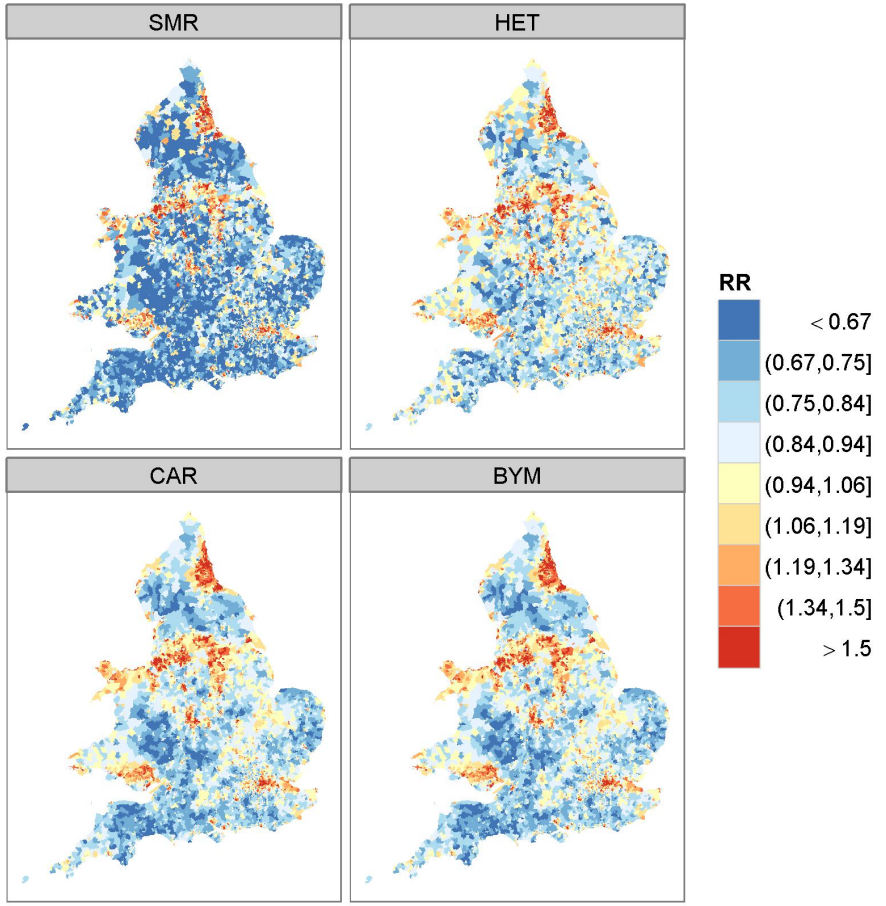
Map of posterior probability of $b_i > 0$

```
> ggplot() + geom_sf(data = out_map,
+                    aes(fill = pp_breaks)) +
+   scale_fill_manual(values = c("blue","orange",
```



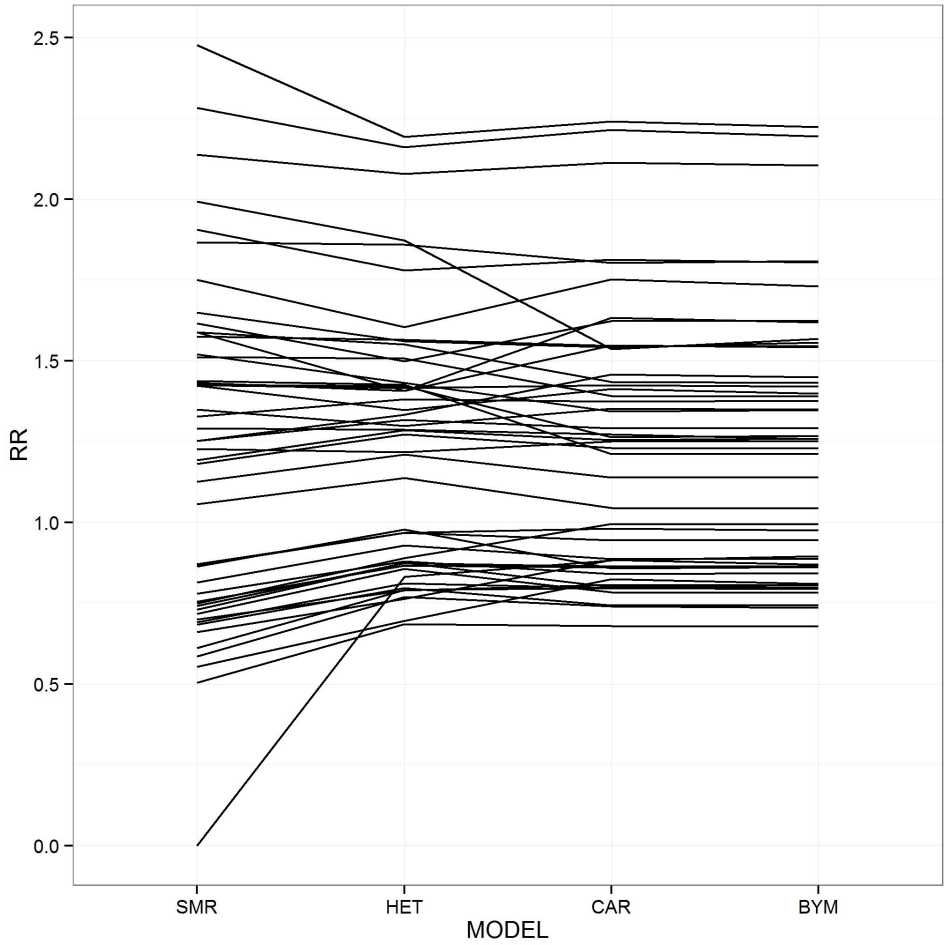
Output from different models

- **SMR** non smoothed RR
- **HET** non spatially smoothed residual RR: $\exp(v)$
- **CAR** spatially smoothed residual RR: $\exp(u)$
- **BYM** spatially and non spatially smoothed residual RR: $\exp(b) = \exp(u + v)$



Output from different models

Comparing maps | **Shrinkage** | Interpretation



- Shrinkage towards the mean due to the **borrowing of strength**

Output from different models

- Smoothed relative risks are more stable (precise than observed)

→ geographical patterns of risk are easier to detect using smoothed maps

- Smoothed relative risks have higher specificity:
  – Possible "false positive" values shrunk towards mean
  – But in danger of over-smoothing (false negatives)

- Visual impact of maps can be very dependent on the choice of colours and cut-points used to shade each region

→ Care must be taken not to over-interpret any patterns identified

Ecological regression with spatial random effects

Disease Mapping vs Ecological Regression

Disease mapping studies

- Focus is on description
- Level of inference is at the aggregate (small area) level

Ecological correlation studies

- Focus is on **explanation**
  - Used for investigating specific exposure-disease hypotheses at small-area scale
  - Poisson regression can be used to model relationship between any area-level exposure measure and incidence/prevalence of disease
  - Such area-level exposure measures include average annual pollution level, proportion of population who smoke, proportion of population living with x km of a landfill site, etc.

Poisson regression with random effects

Straightforward extension of disease mapping model:

Ecological regression with BYM structure

$$
\begin{aligned}
\mathrm{y}_i &\sim \mathrm{Poisson}(E_i \rho_i); \quad i = 1, \ldots, N \\
\log \rho_i &= b_0 + \beta_1 x_i + u_i + v_i \\
\text{residual RR}_i &= \exp(b_i) = \exp(u_i + v_i) \\
\boldsymbol{b} &= \frac{1}{\sqrt{\tau_b}} (\sqrt{1-\phi}\boldsymbol{v}_* + \sqrt{\phi}\boldsymbol{u}_*) \\
v_i &\sim \mathrm{Normal}(0, \sigma_v^2) \quad \mathbf{u} \sim \mathrm{ICAR}(\mathbf{W}, \sigma_u^2)
\end{aligned}
$$

where

- O_i and E_i: Observed and expected nb of cases in each area i
- λ_i: unknown RR
- x area-level covariate of interest
- β_1: parameter associated with the covariate
- $\boldsymbol{v}_*$: standardised version of unstructured random effects, i.i.d.
- $\boldsymbol{u}_*$: standardised version of random effects with spatial structure, conditional distribution

Interpretation of the parameters

- $\exp(\beta_1)$ is the change in risk associated with a unit change in exposure x
- b_i is the random effect in area i
- $\exp(b_i)$ is the residual or adjusted relative risk of disease in area i
  after accounting for the effects of measured covariates and the overall mean risk
- The variance of the random effects reflects the amount of overdispersion in the data (total residual variance = Poisson variance + random effects variance)

Poisson regression with random effects - INLA code

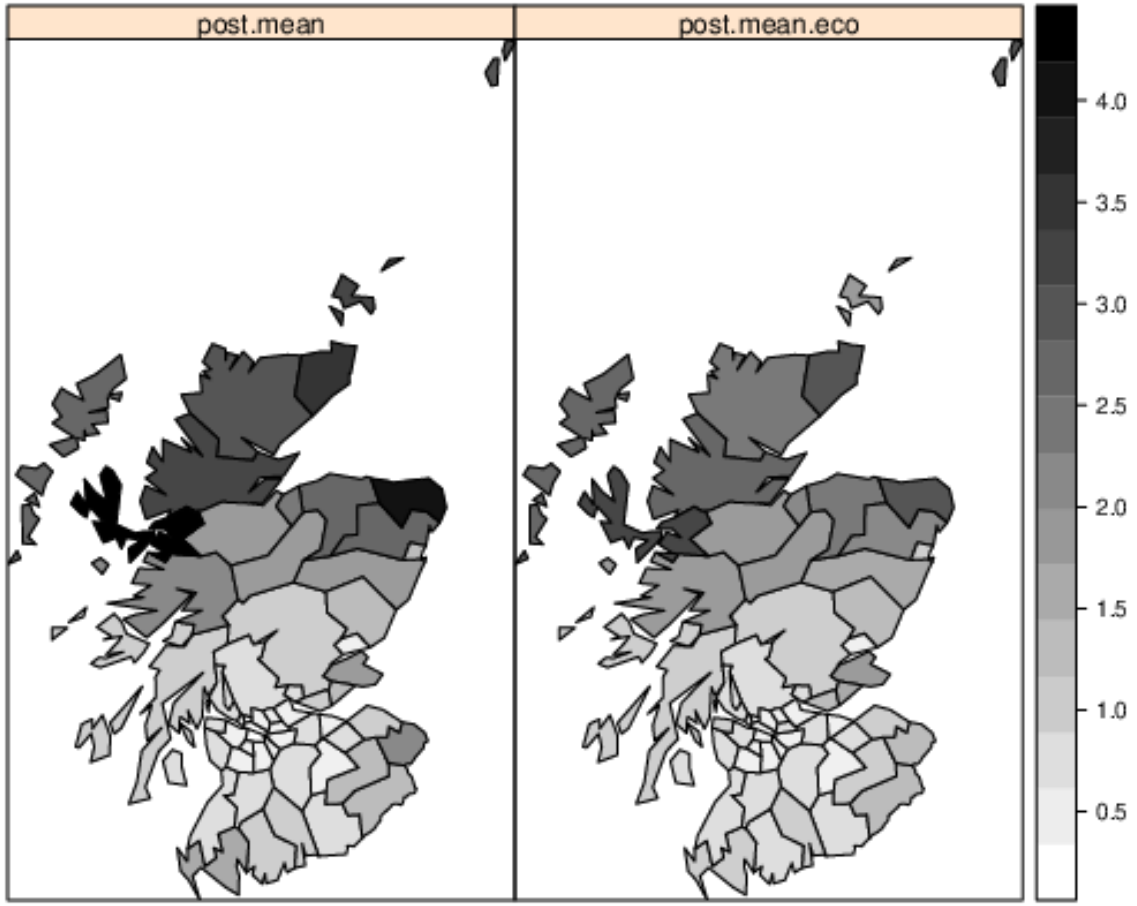
- Continuous covariate

```
> formula.ecoreg.inla <- y ~ 1 + x +
+         f(id,model="bym", graph=graph,
+           hyper=list(prec.spatial=list(
+                 prior="loggamma",param=c(0.01,0.01))))
```

- Categorical covariate

```
> formula.ecoreg.inla <- y ~ 1 + cut(x,breaks=c(0,7,10,24),
+                           include.lowest=TRUE) +
+                           f(id,model="bym", graph=graph,
+                             hyper=list(prec.spatial=list(
+                             prior="loggamma",param=c(0.01,0.01))))
```

Comparison between DM and ecological regression



- Less extreme values when covariates are included

→ part of the spatial variability is explained by the covariates

Poisson regression with random effects

Extension to several variables

$$
\begin{aligned}
y_i &\sim \text{Poisson}(E_i \rho_i); \quad i = 1, \ldots, N \\
\log \rho_i &= b_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + b_i \\
&\ldots
\end{aligned}
$$

- $\exp(\beta_1)$ is the relative risk of disease/death associated with a unit increase in exposure x_1, after adjustment for x_2
- $\exp(\beta_2)$ is the relative risk of disease/death associated with a unit increase in exposure x_2, after adjustment for x_1
- $\exp(b_i)$ is the residual or adjusted relative risk of disease/death in area i after accounting for the effects of measured covariates and the overall mean risk

Summary

- Hierarchical models allow "borrowing of strength" across units
  – posterior distribution of ρ_i for each unit borrows strength from the likelihood contributions for **all** the units, via their joint influence on the posterior estimates of the unknown hyper-parameters

- Judgements of exchangeability need careful assessment
  – units suspected a priori to be systematically different might be modelled by including relevant covariates so that residual variability more plausibly reflects exchangeability
  – subgroups of prior interest should be considered separately

- Mapping geographical variations in disease risk is an important epidemiological technique for suggesting aetiological hypotheses

- When combined with data on geographical variations in exposure, disease mapping techniques can be used to investigate and quantify **ecological** associations between disease risk and potential exposures

References

Besag, J., J. York, and A. Mollie (1991). "Bayesian Image Restoration, with two Applications in Spatial Statistics". In: *Annals of the Institute of Statistical Mathematics* 43, pp. 1-59.

Blangiardo, M. and M. Cameletti (2015). *Spatial and spatio-temporal Bayesian models with R-INLA*. John Wiley & Sons.

Richardson, S., A. Thomson, N. Best, et al. (2004). "Interpreting posterior relative risk estimates in disease-mapping studies". In: *Environmental Health Perspectives* 112.9, pp. 1016-1025.

Simpson, D., H. Rue, A. Riebler, et al. (2017). "Penalising Model Component Complexity: A Principled, Practical Approach to Constructing Priors". In: *Statistical Science* 32.1, pp. 1 - 28. DOI: 10.1214/16-STS576. URL: https://doi.org/10.1214/16-STS576.