

Spatial and Spatio-Temporal Bayesian Models with R-INLA

Marta Blangiardo

Imperial College London

m.blangiardo@imperial.ac.uk

Michela Cameletti

University of Bergamo

michela.cameletti@unibg.it

Last update: Aug. 26, 2019

Spatial epidemiology - definition

Epidemiology: The study of the distribution, causes and control of diseases in human population.

Disease risk depends on the classic epidemiological trio of person (in terms of genetics and behaviour), place and time [spatial epidemiology focuses on 2 and 3].

Place is a surrogate for exposures present at that location, e.g. environmental exposures in water/air/soil, or the lifestyle characteristics of those living in particular areas.

Describing and understanding spatial variation in disease risk and its link with environmental and other potential causes of disease

Types of data

Point-referenced data: the exact location of the case is known
rarely available routinely, can be collected through case-control studies or specialized survey
if location itself is *random*, e.g. measurements of where events occur
⇒ point process statistical framework
if locations are *fixed* (monitoring stations, postcodes in an area) and outcome is measured at each location (e.g. presence/absence of cases, pollution concentrations)
⇒ geostatistical framework.

Area data or count data: locations are areal units with well defined geographical boundaries, usually administrative units
outcome is number of cases aggregated over the area (and time)
most common type of data collected.

We will concentrate on non-infectious diseases and count data.

Need for spatial methods

All epidemiological studies are spatial.

Often the study area is small and/or there is abundant individual-level information and so spatial location is not acting as a surrogate for risk factors.

When are we interested in the spatial component?

Are we explicitly interested in the spatial pattern of disease risk?

Spatial pattern suggests that observations close to each other have more similar values than those far from each other.

→ disease mapping, cluster detection.

Is the clustering a nuisance quantity that we wish to take into account but are not explicitly interested in?

→ spatial regression.

General framework

General framework for small area studies - I

Data for a region of interest/reference area, geographical level and a specific period

England, ward level, 2009-2012

y_i : Observed number of cases in area i

Lung cancer deaths in males aged 45+

Congenital anomalies

n_i : Population at risk in area i

Male population aged 45+

Live births and stillbirths

Parameter of interest Relative risk ϕ_i in each area compared with the chosen reference area.

General framework for small area studies - II

Standard statistical model if rare disease and/or small areas

$$y_i \sim \text{Poisson}(\phi_i E_i)$$

where E_i = expected nb of cases in area i

ϕ_i estimated by Standardised Mortality/Incidence Ratio (maximum likelihood estimator)

$$\hat{\phi}_i = \text{SMR}_i \text{ or } \text{SIR}_i = \frac{y_i}{E_i} \quad \text{and} \quad \text{Var}(\hat{\phi}_i) = \frac{\phi_i}{E_i} \rightarrow \text{Var}(\hat{\phi}) = \frac{y_i}{E_i^2}$$

Recall: $X \sim \text{Poisson}(\mu) \Leftrightarrow \text{E}(X) = \text{Var}(X) = \mu$

Expected numbers of cases - definition

Expected nb of cases if the population had the same stratum-specific mortality/incidence rates as in a reference area.

Adjustments (strata): age, gender, maternal age,...

Indirect standardisation

$$E_i = \sum_j n_{ij} r_j$$

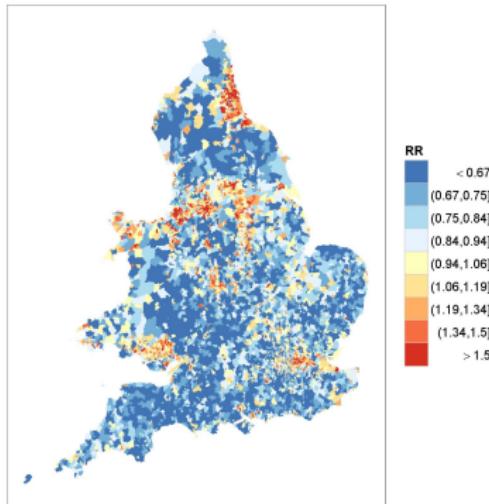
with

r_j : disease rate for stratum j in the reference population

n_{ij} : population at risk in area i , stratum j

If internal comparison: $\sum_{i=1}^N y_i = \sum_{i=1}^N E_i$

Lung cancer incidence in males, 1985-2009, England and Wales



SIRs at ward level

	Min	Q1	Median	Q3	Max
y	0	26	47	84	456
E	3.25	32.14	53.60	82.47	390.49
SMR	0	0.70	0.89	1.13	2.63

Is the variability real or simply reflecting unequal E_i s?

Have the highlighted areas truly a raised relative risk?

Problems with mapping SMRs

Common practice is to map SMRs

- SMR very imprecise for rare diseases and/or areas with small populations
- SMR in each area is estimated independently
 - makes no use of risk estimates in other areas of the map, even though these are likely to be similar.
- ⇒ Highlights extreme risk estimates based on small numbers.
- ⇒ Ignores possible spatial correlation between disease risk in nearby areas due to possible dependence on spatially varying risk factors.

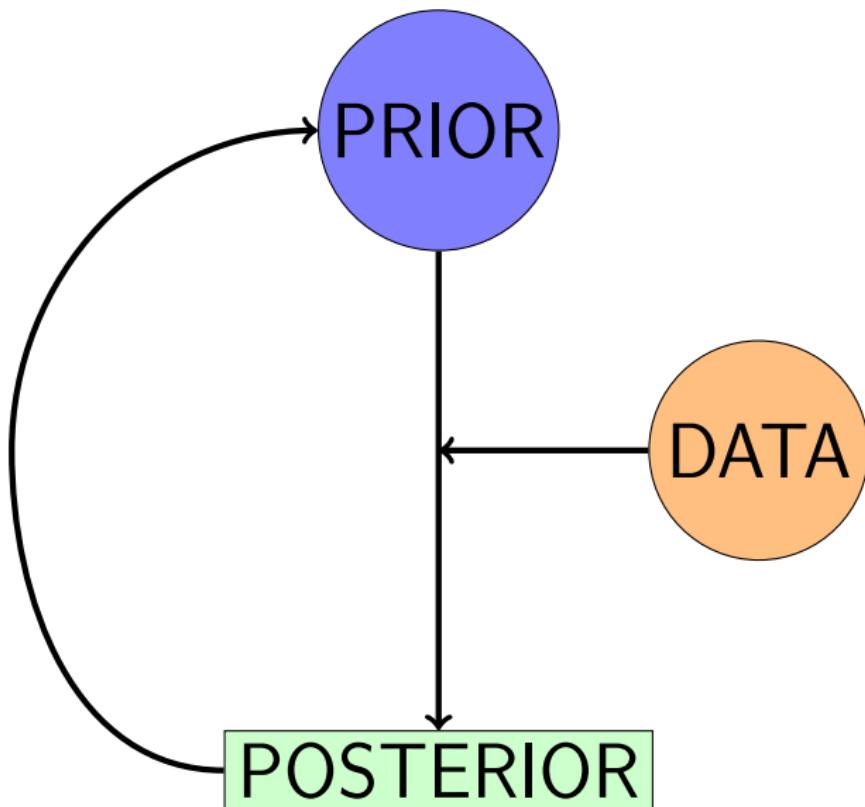
Problems addressed using Bayesian 'smoothing' estimators in a hierarchical formulation:

Poisson-logNormal model: non spatial smoothing

Poisson-logNormal-spatial model: spatial and non spatial smoothing

Bayesian inference

Bayesian thinking



For more details see e.g. [Gelman et al., 2014], [Hoff, 2009].

Bayesian inference

Makes fundamental distinction between

Observable quantities y , i.e. the data

Unknown quantities θ

θ can be statistical parameters, missing data, mismeasured data, ...

→ parameters are treated as random variables

→ in the Bayesian framework, we make probability statements about model parameters

! in the Frequentist framework, parameters are fixed non-random quantities and the probability statements concern the data

As with any statistical analysis, we start building a model which specifies $p(y | \theta)$

This is the **sampling distribution**, which relates all variables into a '**full probability model**'.

Bayesian inference [continued]

From a Bayesian point of view

θ is unknown so should have a **probability distribution** reflecting our uncertainty about it before seeing the data

→ need to specify a **prior distribution** $p(\theta)$

y is known so we should condition on it

→ use **Bayes theorem** to obtain conditional probability distribution for unobserved quantities of interest given the data:

$$p(\theta | y) = \frac{p(y | \theta) p(\theta)}{p(y)} \propto p(y | \theta) p(\theta)$$

This is the **posterior distribution**.

The prior distribution $p(\theta)$, expresses our uncertainty about θ **before** seeing the data.

The posterior distribution $p(\theta | y)$, expresses our uncertainty about θ **after** seeing the data.

Bayesian inference: the posterior distribution

Posterior distribution forms basis for all inference — can be summarised to provide:

- point and interval estimates of Quantity of Interest (QOI), e.g. treatment effect, small area estimates, ...
- point and interval estimates of any function of the parameters
- probability that QOI exceeds a critical threshold
- prediction of QOI in a new unit
- prior information for future experiments, trials, surveys, ...
- inputs for decision making
- ...

Computation

Why is computation important?

Bayesian inference centres around the posterior distribution

$$p(\boldsymbol{\theta}|\mathbf{y}) \propto p(\mathbf{y}|\boldsymbol{\theta}) \times p(\boldsymbol{\theta})$$

where $\boldsymbol{\theta}$ is typically a large vector of parameters $\boldsymbol{\theta} = \{\theta_1, \theta_2, \dots, \theta_k\}$

$p(\mathbf{y}|\boldsymbol{\theta})$ and $p(\boldsymbol{\theta})$ will often be available in closed form, but $p(\boldsymbol{\theta}|\mathbf{y})$ is usually not analytically tractable, and we want to

obtain marginal posterior

$$p(\theta_i|\mathbf{y}) = \int p(\boldsymbol{\theta}|\mathbf{y}) d\boldsymbol{\theta}_{(-i)}$$

where $\boldsymbol{\theta}_{(-i)}$ denotes the vector of θ s excluding θ_i

calculate properties of $p(\theta_i|\mathbf{y})$, such as mean ($= \int \theta_i p(\theta_i|\mathbf{y}) d\theta_i$), tail areas ($= \int_T^\infty p(\theta_i|\mathbf{y}) d\theta_i$) etc.

Approaches to computation

Simulation-based methods

Markov Chain Monte Carlo (MCMC): class of algorithm to sample from the posterior distribution (when not available in a closed form)

MCMC methods are flexible and able to deal with virtually any type of data and model, but they involve computationally- and time- intensive simulations to obtain the posterior distribution for the parameters. For this reason the complexity of the model and the database dimension often remain fundamental issues.

Software: BUGS, JAGS, STAN, MCMCglmm, CARBayes R packages,...

Approximate methods

The Integrated Nested Laplace Approximation (INLA) algorithm proposed by [Rue et al., 2009] is a *deterministic* algorithm for Bayesian inference.

The INLA algorithm is designed for the class of *latent Gaussian models* and compared to MCMC it provides (as) accurate results in a shorter time.

INLA has become very popular amongst statisticians and applied researchers and in the past few years the number of papers reporting usage and extensions of the INLA method has increased considerably (see [Rue et al., 2017]).

<http://www.r-inla.org/>

The website contains source code, examples, papers and reports discussing the theory and applications of INLA.

There is also a discussion forum where users can post queries and requests of help.

Information about how to install the R-INLA package (which is not available from the CRAN) are provided at <http://www.r-inla.org/download>

Latent Gaussian models (LGMs)

Latent Gaussian models (LGMs)

The general problem of (parametric) inference is posited by assuming a probability model for the observed data $\mathbf{y} = (y_1, \dots, y_n)$, as a function of some relevant parameters

$$\mathbf{y} | \boldsymbol{\theta}, \boldsymbol{\psi} \sim p(\mathbf{y} | \boldsymbol{\theta}, \boldsymbol{\psi}) = \prod_{i=1}^n p(y_i | \boldsymbol{\theta}, \boldsymbol{\psi})$$

Often (in fact for a surprisingly large range of models!), we can assume that the parameters are described by a **Gaussian Markov Random Field** (GMRF, [Rue and Held, 2005])

$$\begin{aligned}\boldsymbol{\theta} | \boldsymbol{\psi} &\sim \text{Normal}(\mathbf{0}, \mathbf{Q}^{-1}(\boldsymbol{\psi})) \\ \theta_i \perp\!\!\!\perp \theta_j | \boldsymbol{\theta}_{-i,j} &\iff Q_{ij}(\boldsymbol{\psi}) = 0\end{aligned}$$

where

The precision matrix \mathbf{Q} depends on some **hyperparameters** $\boldsymbol{\psi}$.

The notation “ $-i, j$ ” indicates all the other elements of the parameters vector, excluding elements i and j .

The components of $\boldsymbol{\theta}$ are supposed to be *conditionally independent* (given the hyperparameters) with the consequence that \mathbf{Q} is a sparse precision matrix.

LGMs as a general framework

In general

$$\begin{aligned}\mathbf{y} \mid \boldsymbol{\theta}, \boldsymbol{\psi} &\sim \prod_i p(y_i \mid \boldsymbol{\theta}, \boldsymbol{\psi}) && \text{("data model")} \\ \boldsymbol{\theta} \mid \boldsymbol{\psi} &\sim p(\boldsymbol{\theta} \mid \boldsymbol{\psi}) = \text{Normal}(0, \mathbf{Q}^{-1}(\boldsymbol{\psi})) && \text{("GMRF prior")} \\ \boldsymbol{\psi} &\sim p(\boldsymbol{\psi}) && \text{("hyperprior")}\end{aligned}$$

The dimension of $\boldsymbol{\theta}$ can be very large (eg 10^2 - 10^5).

Conversely, the dimension of $\boldsymbol{\psi}$ must be relatively small (less than 20 is recommended) to avoid an exponential increase in the computational costs of the model.

LGMs as a general framework

A very general way of specifying the problem is specifying a distribution for y_i characterized by a parameter ϕ_i defined (on a suitable scale) as a function of a structured additive predictor η_i , such that $g(\phi_i) = \eta_i$ (e.g. logarithm for Poisson data):

$$\eta_i = \beta_0 + \sum_{m=1}^M \beta_m x_{mi} + \sum_{l=1}^L f_l(z_{li})$$

where

β_0 is the intercept

$\boldsymbol{\beta} = \{\beta_1, \dots, \beta_M\}$ quantify the effect of the covariates $\mathbf{x} = (\mathbf{x}_1, \dots, \mathbf{x}_M)$ on the response

$\mathbf{f} = \{f_1(\cdot), \dots, f_L(\cdot)\}$ is a set of functions defined in terms of some covariates $\mathbf{z} = (\mathbf{z}_1, \dots, \mathbf{z}_L)$

and then assume

$$\boldsymbol{\theta} = \{\beta_0, \boldsymbol{\beta}, \mathbf{f}\} \sim \text{Normal}(\mathbf{0}, \mathbf{Q}^{-1}(\psi)) = \text{GMRF}(\psi)$$

NB: This of course implies some form of Normally-distributed marginals for β_0 , $\boldsymbol{\beta}$ and \mathbf{f} .

LGMs as a general framework — examples

Upon varying the form of the functions $f_l(\cdot)$, this formulation can accommodate a wide range of models (see [Martins et al., 2013] for a review)

Standard regression

$$f_l(\cdot) = \text{NULL}$$

Hierarchical models

$$f_l(\cdot) \sim \text{Normal}(0, \sigma_f^2)$$

$$\sigma_f^2 \sim \text{some common distribution}$$

Spatial and spatio-temporal models

Areal data: $f_1(\cdot) \sim \text{CAR}$ (Spatially structured effects)

$$f_2(\cdot) \sim \text{Normal}(0, \sigma_{f_2}^2)$$
 (Unstructured residual)

Geostatistical data: $f(\cdot) \sim \text{Gaussian field}$

Temporal component: $f(\cdot) \sim \text{RW}$

Spline smoothing

$$f_l(\cdot) \sim \text{AR}(\rho, \sigma_\varepsilon^2)$$

Survival models, logGaussian Cox Processes, etc.

Back to disease mapping: Smoothed estimates of ϕ_i

In the context of the disease mapping:

Poisson-logNormal model

$$\begin{aligned}y_i &\sim \text{Poisson}(\phi_i E_i) \\ \eta_i &= \log \phi_i = \beta_0 + v_i\end{aligned}$$

where $v_i = f(z_i)$ and z_i is the area ID.

So

$$\begin{aligned}v_i &\sim \text{Normal}(0, \sigma_v^2) \\ \beta_0 &\sim \text{Normal}(0, 10^6)\end{aligned}$$

As seen in the previous slides

Parameters: $\theta = \{\beta_0, v\}$

Hyperparameters: $\psi = \{\sigma_v^2\}$

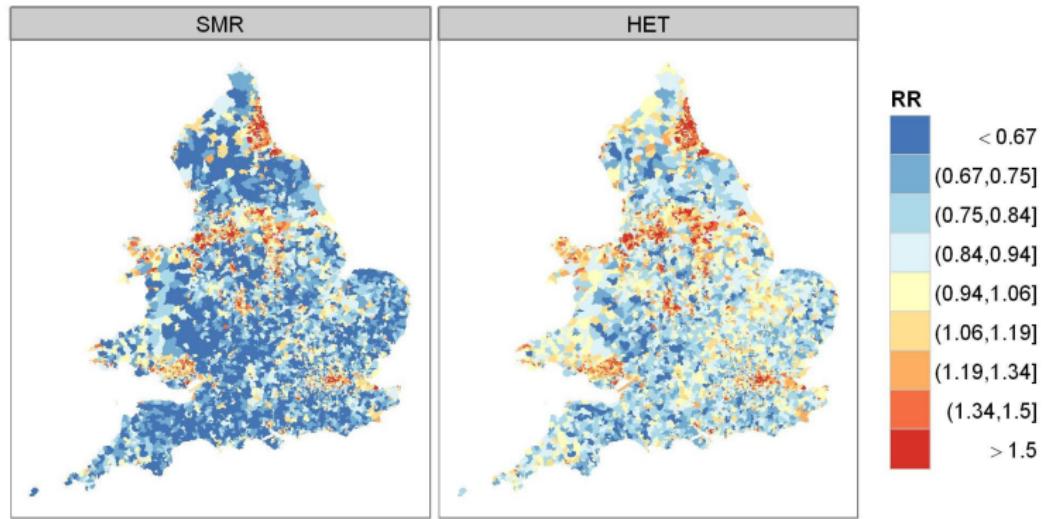
Note that

y_i, E_i : observed and expected nb of cases in area i (known data)

$\phi_i = \exp(\beta_0 + v_i)$: unknown RR in area i compared with expected risk based on age and sex of population (reference area)

v_i : area-specific random effects to take into account overdispersion.

Mapping smoothed vs raw estimates of ϕ



Shrinkage towards the mean when using a hierarchical model.
Higher specificity, but danger of over smoothing.
Important to keep the same cut points.
Careful to not over-interpret the maps.

The INLA approach

First ingredient

The first “ingredient” of the INLA approach is the definition of **conditional probability**, which holds for any pair of variables (x, z) — and, technically, provided $p(z) > 0$

$$p(x | z) =: \frac{p(x, z)}{p(z)} \rightarrow p(x, z) = p(x | z)p(z)$$

$p(x | z)$ can be re-written as

$$p(z) = \frac{p(x, z)}{p(x | z)}$$

In particular, a conditional version can be obtained further considering a third variable w as

$$p(z | w) = \frac{p(x, z | w)}{p(x | z, w)}$$

which is particularly relevant to the Bayesian case.

Second ingredient: Laplace Approximation (LA)

Approximate an integral $\int f(x)dx = \int \exp(\log(f(x)))dx$ where $f(x)$ is the density function of a random variable X .

Main idea: Represent $\log(f(x))$ by means of a **Taylor's series expansion** evaluated in the mode x^* of X .

This leads to $f(x) \approx \text{Normal}(x^*, \sigma^{2*})$, where $\sigma^{2*} = -1/\left.\frac{\partial^2 \log f(x)}{\partial x^2}\right|_{x=x^*}$

Laplace approximation — example

As an example we consider the case of the Gamma distribution with density function

$$f(x) = \frac{b^a}{\Gamma(a)} \exp(-bx) x^{a-1} \quad x, a, b > 0$$

For computing the Laplace approximation we need the following quantities:

$$l(x) = \log f(x) = (a-1) \log x - bx + \text{constant}$$

$$l'(x) = \frac{\partial \log f(x)}{\partial x} = \frac{a-1}{x} - b$$

$$l''(x) = \frac{\partial^2 \log f(x)}{\partial x^2} = -\frac{a-1}{x^2}$$

Then:

Solving $l'(x) = 0$ we obtain the mode $x^* = \frac{a-1}{b}$ (for $a > 1$).

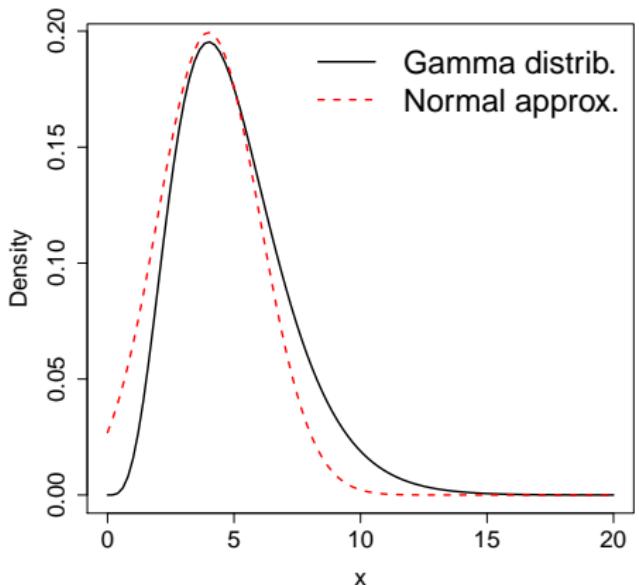
Evaluating $-\frac{1}{l''(x)}$ at the mode gives $\sigma^2* = \frac{a-1}{b^2}$

Thus, the Laplace approximation of the Gamma distribution is

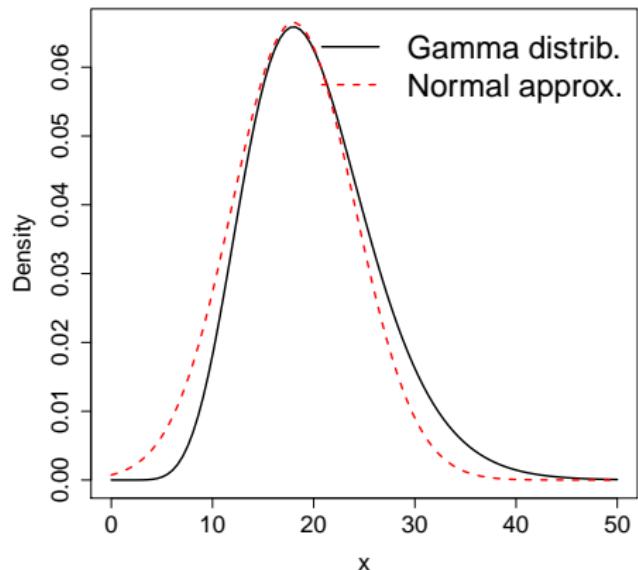
$$\text{Gamma}(a, b) \approx \text{Normal}\left(x^* = \frac{a-1}{b}, \sigma^2* = \frac{a-1}{b^2}\right).$$

Laplace approximation — example

$\text{Gamma}(a = 5, b = 1)$



$\text{Gamma}(a = 10, b = 5)$



Exact and approximated values of $\int_{\alpha}^{\beta} f(x)dx$ where $f(x)$ is the density function of the $\text{Gamma}(a, b)$ distribution.

(a, b)	(α, β)	Exact value	Approximated value (Laplace)
$(5, 1)$	$(2, 6)$	0.6622905	0.6686424
$(5, 1)$	$(1, 10)$	0.9670875	0.9126693
$(10, 0.5)$	$(18, 22)$	0.2468976	0.2452272
$(10, 0.5)$	$(10, 40)$	0.9631765	0.9002946

Integrated Nested Laplace Approximation (INLA)

In a Bayesian LGM, the required distributions are

$$\begin{aligned} p(\theta_i | \mathbf{y}) &= \int p(\theta_i, \psi | \mathbf{y}) d\psi = \int \color{red}{p(\psi | \mathbf{y})} \color{orange}{p(\theta_i | \psi, \mathbf{y})} d\psi \\ p(\psi_k | \mathbf{y}) &= \int \color{red}{p(\psi | \mathbf{y})} d\psi_{-k} \end{aligned}$$

Estimate (approximate)

$$p(\psi | \mathbf{y}) = \frac{p(\theta, \psi | \mathbf{y})}{p(\theta | \psi, \mathbf{y})} \approx \left. \frac{p(\psi)p(\theta | \psi)p(\mathbf{y} | \theta, \psi)}{\tilde{p}(\theta | \psi, \mathbf{y})} \right|_{\theta=\hat{\theta}(\psi)} =: \tilde{p}(\psi | \mathbf{y})$$

$$\begin{aligned} p(\theta_i | \psi, \mathbf{y}) &= \frac{p(\{\theta_i, \theta_{-i}\} | \psi, \mathbf{y})}{p(\theta_{-i} | \theta_j, \psi, \mathbf{y})} \approx \\ &\quad \left. \frac{p(\psi)p(\theta | \psi)p(\mathbf{y} | \theta, \psi)}{\tilde{p}(\theta_{-i} | \theta_i, \psi, \mathbf{y})} \right|_{\theta_{-i}=\hat{\theta}_{-i}(\theta_i, \psi)} =: \tilde{p}(\theta_i | \psi, \mathbf{y}) \end{aligned}$$

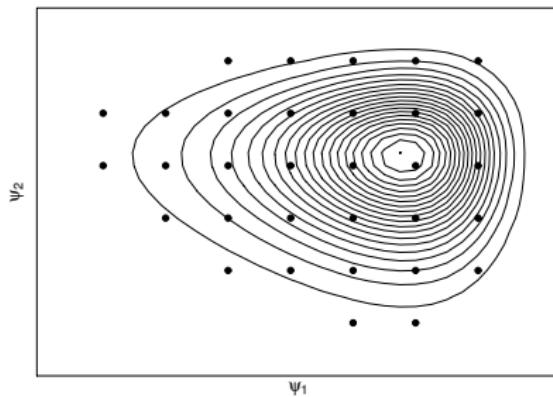
where \tilde{p} indicates the Laplace approximation and $\hat{\theta}$ is the mode

Use numerical integration to obtain the marginals

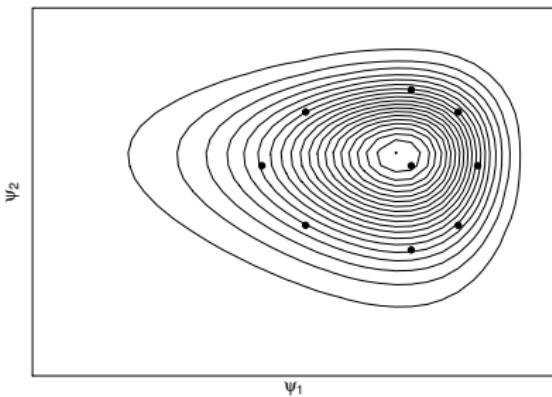
Integrated Nested Laplace Approximation (INLA)

Operationally, the INLA algorithm proceeds with the following steps:

1. Explore the joint posterior for the hyperparameters $\tilde{p}(\psi | \mathbf{y})$ and produce a grid of “good” integration points $\{\psi^{(j)}\}$ associated with the bulk of the mass, together with a corresponding set of area weights $\{\Delta_j\}$:



Grid strategy



Central Composite Design strategy (CCD)

The CCD strategy is the default one in R-INLA: it produces a lower number of points which are however enough to capture the variability of the joint distribution (see [Martins et al., 2013]).

Integrated Nested Laplace Approximation (INLA)

Recall that the required distributions are

$$\begin{aligned} p(\theta_i | \mathbf{y}) &\approx \int \tilde{p}(\boldsymbol{\psi} | \mathbf{y}) \tilde{p}(\theta_i | \boldsymbol{\psi}, \mathbf{y}) d\boldsymbol{\psi} \\ p(\psi_k | \mathbf{y}) &\approx \int \tilde{p}(\boldsymbol{\psi} | \mathbf{y}) d\boldsymbol{\psi}_{-k} \end{aligned}$$

After the grid exploration:

1. Obtain the marginal posterior $\tilde{p}(\psi_k | \mathbf{y})$ using an interpolation algorithm based on the values of the density $\tilde{p}(\boldsymbol{\psi} | \mathbf{y})$ evaluated in the integration points $\{\boldsymbol{\psi}^{(j)}\}$
2. Obtain the marginal posteriors $\tilde{p}(\theta_i | \mathbf{y})$ using **numerical integration**

$$\tilde{p}(\theta_i | \mathbf{y}) \approx \sum_j \tilde{p}(\boldsymbol{\psi}^{(j)} | \mathbf{y}) \tilde{p}(\theta_i | \boldsymbol{\psi}^{(j)}, \mathbf{y}) \Delta_j$$

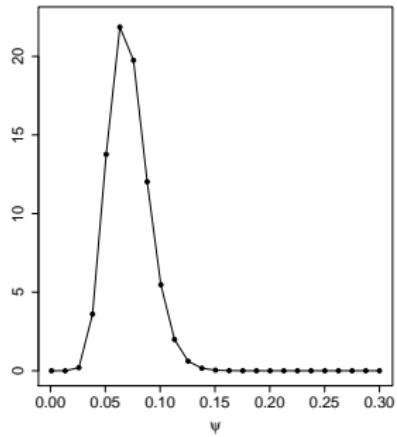
INLA in practice

$$y_i \mid \theta, \psi \sim \text{Normal}(\eta_i, \sigma^2)$$

where $\eta_i = \theta = \mu$ and $\psi = 1/\sigma^2$. A $\text{Gamma}(a, b)$ is used as a prior for ψ .

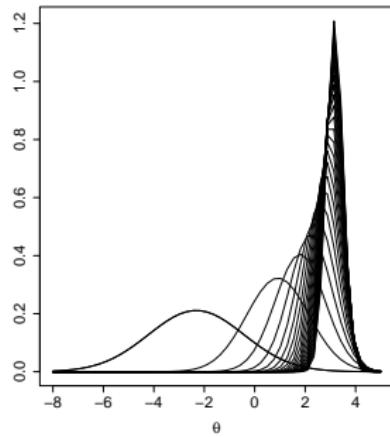
$$\tilde{p}(\psi \mid \mathbf{y})$$

• = grid values $\{\psi^{(j)}\}$

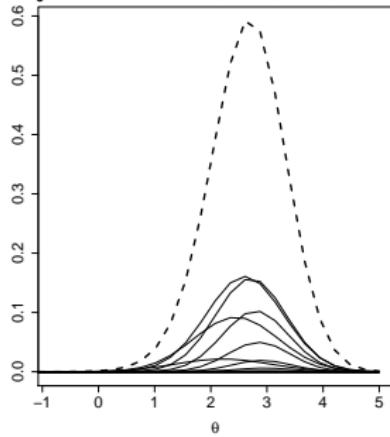


$$\tilde{p}(\theta \mid \psi^{(j)}, \mathbf{y})$$

for each ψ in $\{\psi^{(j)}\}$



$$\tilde{p}(\theta \mid \mathbf{y}) =$$
$$\sum_j \tilde{p}(\theta \mid \psi^{(j)}, \mathbf{y}) \tilde{p}(\psi^{(j)}) \Delta_j$$



Summary so far

The INLA approach is not a rival/competitor/replacement to/of MCMC, just a better option for the class of LGMs.

The basic idea behind the INLA procedure is simple:

- Repeatedly use Laplace approximation and take advantage of computational simplifications due to the structure of the model;
- Use numerical integration to compute the required posterior marginal distributions.

Complications are mostly computational and occur when:

- Extending to a large number of hyperparameters;
- Markedly non-Gaussian observations.

Adding a spatial structure

Why adding a spatial structure?

Poisson-logNormal model based on the assumption that the observations in the data set are identically distributed and independent.

⇒ Independence makes much of the mathematics tractable.

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.

A conditional spatial model

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.

Intrinsic CAR model

Common definition [Besag, 1974]

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

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

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

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)

Poisson model with BYM random effects I

Besag, York and Mollie (BYM, [Besag et al., 1991]) recommend combining the ICAR prior and the standard normal prior to allow for both

- spatially unstructured latent covariates \mathbf{v} modelled as iid
- ⇒ global smoothing
- spatially correlated latent covariates \mathbf{u} modelled as ICAR
- ⇒ local smoothing

Convolution or BYM model

$$\begin{aligned}y_i &\sim \text{Poisson}(\phi_i E_i) \\ \log \phi_i &= \beta_0 + v_i + u_i \\ v_i &\sim \text{Normal}(0, \sigma_v^2) \\ \mathbf{u} &\sim \text{ICAR}(\mathbf{W}, \sigma_u^2)\end{aligned}$$

Priors (vague, non informative): $\sigma_v^2, \sigma_u^2, \beta_0$

Quantification of spatial and unstructured variability in the BYM model

σ_v^2 reflects variability of the unstructured REs between areas

σ_u^2 reflects variability of the spatial effect

Relative contribution of spatial vs. unstructured heterogeneity:

$$\text{frac}_{\text{spatial}} = \sigma_u^2 / (\sigma_u^2 + \sigma_v^2)$$

$\text{frac}_{\text{spatial}} \rightarrow 1 \Rightarrow$ spatial heterogeneity dominates

$\text{frac}_{\text{spatial}} \rightarrow 0 \Rightarrow$ unstructured heterogeneity dominates

Use the BYM model with care!

ICAR model defined above is improper: the overall mean of the u_i is not defined. So additional constraints need to be imposed:

sum-to-zero constraint: $\sum_i u_i = 0$ (implemented in INLA)

Inference may be sensitive to choice of hyperprior for the random effects variance or precision parameters (σ_u^2 and σ_v^2)

Interpretation of the variance parameters (conditional, marginal)

Total variation of the log residual $RR = \sigma_v^2 + \sigma_u^2$

Poisson model with BYM random effects II

Choice of the adjacency matrix (neighbours) → 2 areas are adjacent if common borders

$\text{RR}_i = \exp(\beta_0 + v_i + u_i)$: RR in area i relative to the age/sex structure (used to estimate the E_i)

residual $\text{RR}_i = \exp(v_i + u_i)$: residual RR in area i relative to the region average after adjusting for the overall risk

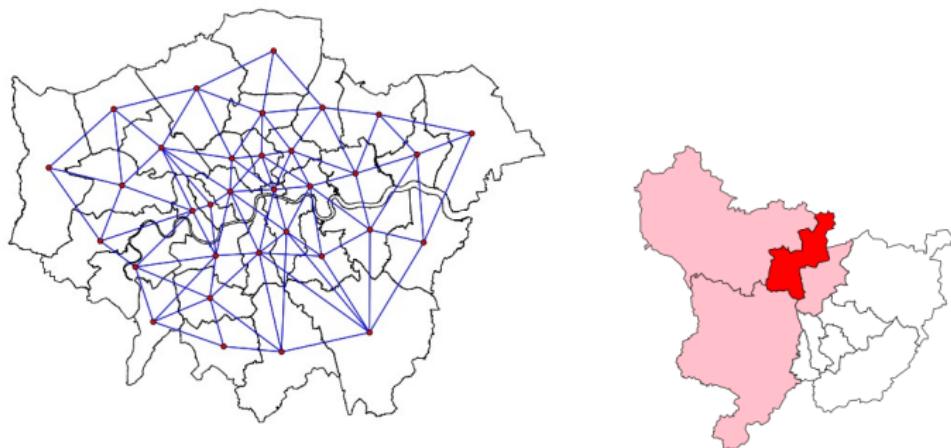
Neighbourhood definition

Common approach: 2 areas are neighbours if they share a common border (or point)

→ Adjacency matrix implemented in INLA

An area cannot be specified as its own neighbour

Adjacency matrix must be symmetric



Lung cancer incidence in males, 1985-2009, England and Wales

Here we replace the unstructured Normal random effects prior for the log relative risks by a convolution (ICAR + unstructured Normal) prior:

$$\begin{aligned}y_i &\sim \text{Poisson}(\phi_i E_i) \\ \log \phi_i &= \beta_0 + v_i + u_i \\ v_i &\sim \text{Normal}(0, \sigma_v^2) \\ \mathbf{u} &\sim \text{ICAR}(\mathbf{W}, \sigma_u^2)\end{aligned}$$

Data: \mathbf{y} and \mathbf{E} , observed and expected cases, \mathbf{W} adjacency matrix

Priors: $\sigma_v^2, \sigma_u^2, \beta_0$

Parameters of interest:

$$\text{RR} = \exp(\beta_0 + v_i + u_i)$$

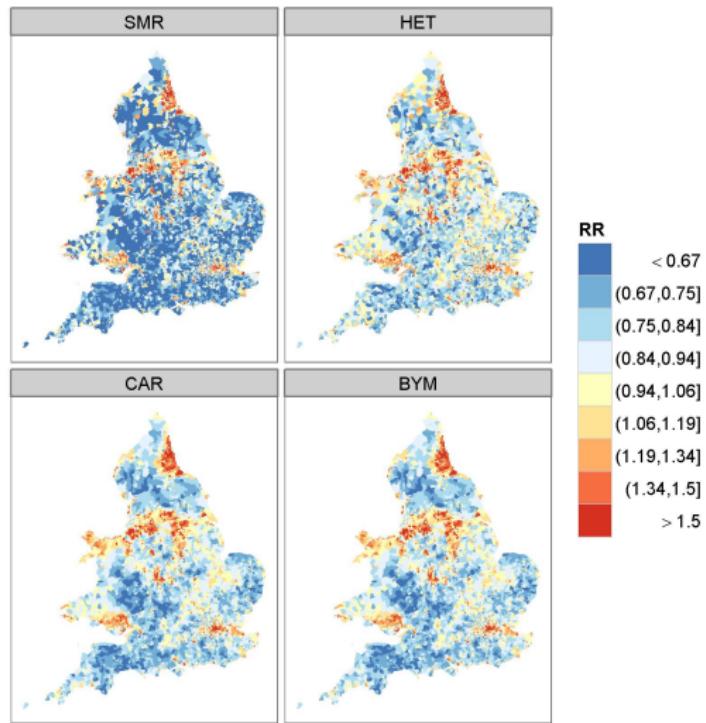
Residual RR of lung cancer incidence in males, 1985-2009, England and Wales II

SMR non smoothed RR

HET non spatially smoothed RR

ICAR spatially smoothed RR

BYM spatially and non spatially smoothed RR



Other conditional autoregressive structures

BYM is not the only option to account for spatial correlation at small area.

A popular alternative consists in the modelling proposed by Leroux ([Leroux et al., 2000]).

Leroux model

$$u_i \mid u_j, j \neq i \sim \text{Normal} \left(\frac{\rho \sum_{j \in \partial_i} u_j + (1 - \rho)\mu}{n_i\rho + 1 - \rho}, \frac{\sigma_u^2}{n_i\rho + 1 - \rho} \right)$$

The additional parameter ρ governs the amount of spatial autocorrelation. As μ is included in the formulation there is no need for an intercept in the log-linear specification, i.e. $\log \phi_i = u_i$.

Priors: σ_u^2, μ, ρ .

Not directly available in INLA.

Posterior Probability

Mapping the posterior mean relative risk does not make full use of the output of the Bayesian analysis that provides, for each area, samples from the whole posterior distribution of the relative risk.

Mapping the probability that a relative risk is greater than a specified threshold of interest has been proposed by several authors (e.g. [Clayton and Bernardinelli, 1996]).

Very effective method to identify areas characterised by elevated risk.

How to classify areas as having elevated risk

We define the decision rule $D(c, RR_0)$, which depends

→ on a cutoff probability c

→ a reference threshold RR_0 .

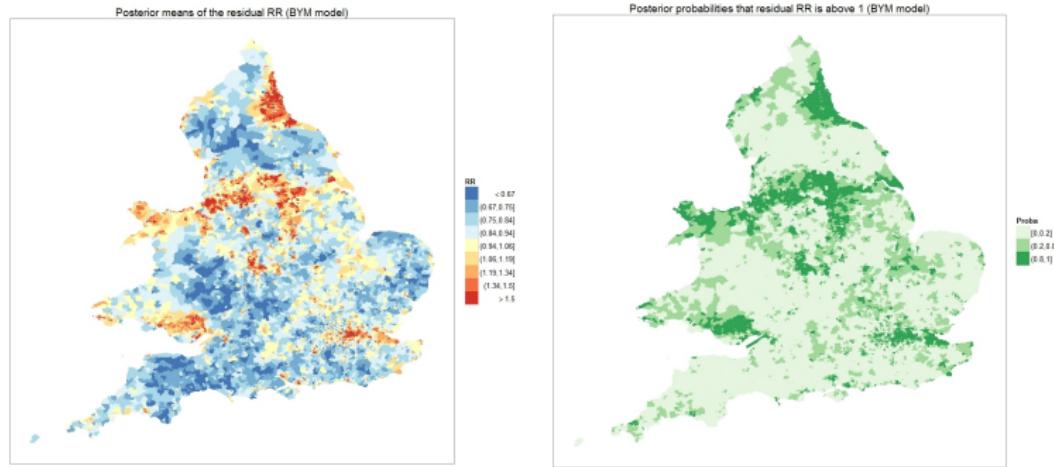
Area i is classified as having an elevated risk according to
 $D(c, RR_0) \leftrightarrow \text{Prob}(RR_i > RR_0) > c$.

[Richardson et al., 2004] presented a simulation study to find the best parameters c and RR_0 and proposed $c = 0.8$ and $RR_0 = 1$.

Posterior probabilities of interest

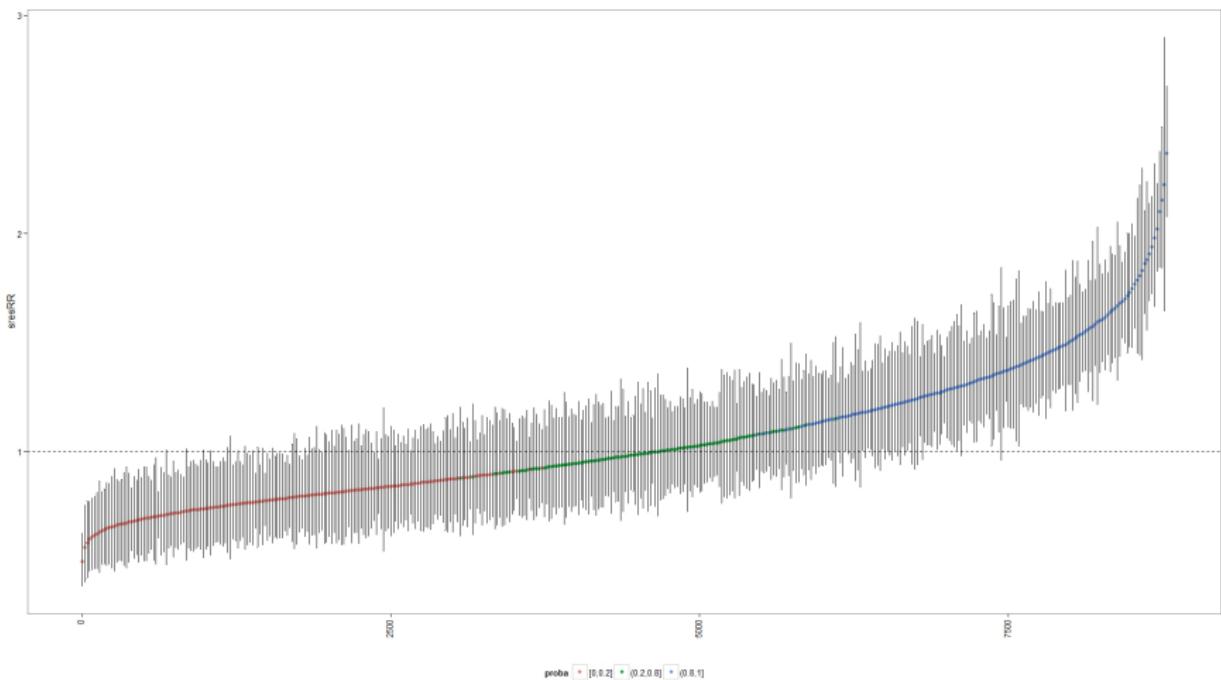
$$\text{Prob}(e^{v_i+u_i} > 1) = \text{Prob}(e^{\beta_0+v_i+u_i} > e^{\beta_0}) = \text{Prob}(RR_i > e^{\beta_0}).$$

Lung cancer incidence in males, 1985-2009, England and Wales



Map of the smoothed residual RRs and posterior probabilities that the RR is above the average risk

Lung cancer incidence in males, 1985-2009, England and Wales



From space to space-time

Disease mapping: Extending space to space-time

Disease mapping is usually carried out on aggregated data over a time period.

Rather than suppressing the time dimension, it can be interesting to use models that combine the space and time dimension.

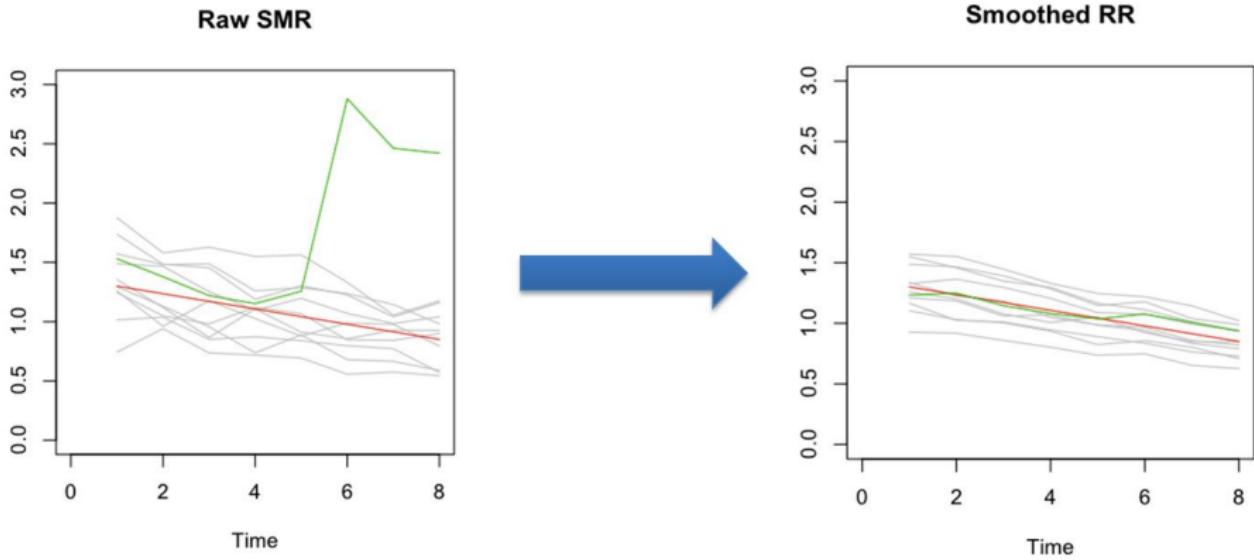
The stability (or not) of the spatial pattern can aid interpretation.

The specific space-time components of the model can potentially pinpoint unusual/emerging hazards.

Data y_{it} and E_{it} : the observed and expected number of cases in area i at time t

$$E_{it} = \sum_k n_{itk} r_k, \text{ where } r_k \text{ reference rate for stratum (age, gender,...)}$$

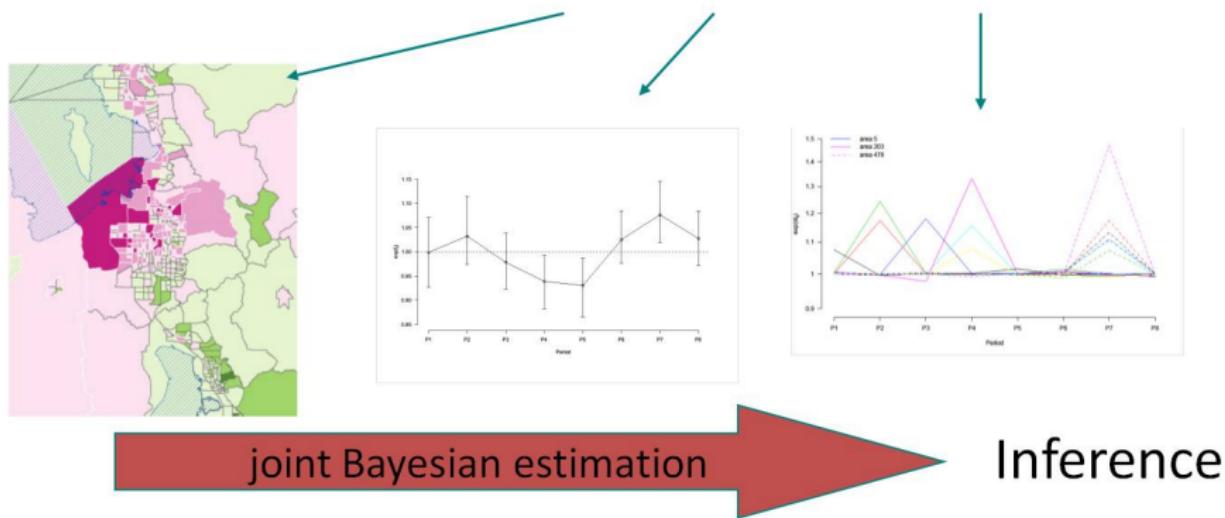
Schematic representation I



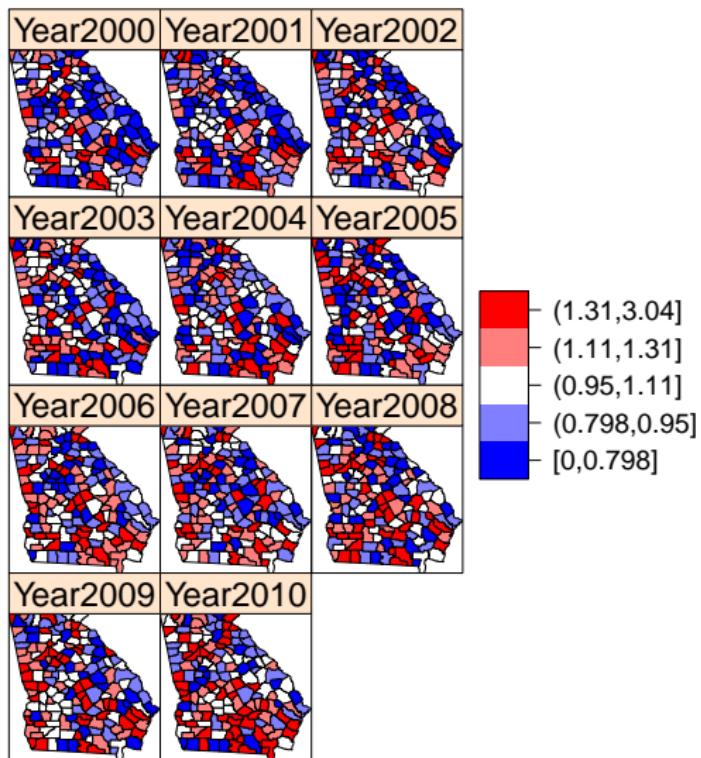
Schematic representation II

Noise model: Poisson/Binomial

Latent structure: Space + Time + Interactions



Georgia low birth weight - temporal SMRs

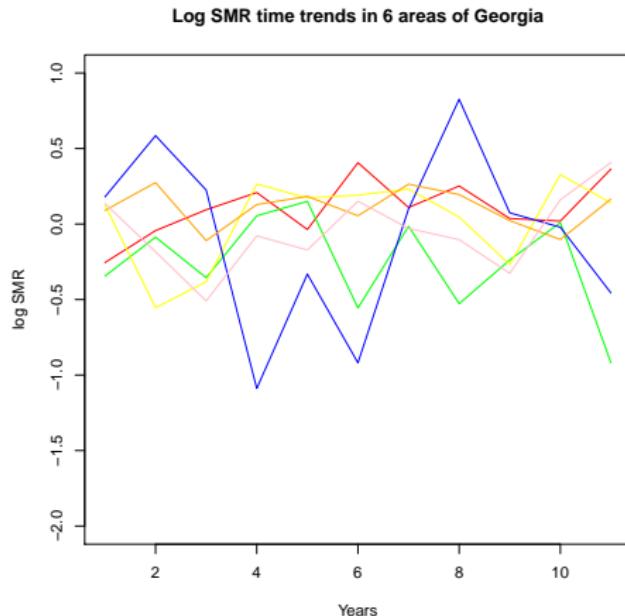


Low birth weight babies ($< 2500\text{g}$) born between 2000 and 2010 in Georgia (county level).

County specific yearly SMR

Georgia low birth weight - temporal SMRs for selected counties

Log SMR time trends in 6 counties of Georgia



"Messy" trends - different variability for different areas

Spatio-temporal model

$$\begin{aligned}y_{it} &\sim \text{Poisson}(\phi_{it} E_{it}) \\ \log \phi_{it} &= \beta_0 + u_i + v_i + ?\end{aligned}$$

where

β_0 overall log RR in Georgia over the 11-year period;

$v_i \sim \text{Normal}(0, \sigma_v^2)$ spatially unstructured RE;

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

How do we model time?

Contrary to spatial models, there is a natural order to any time series data which is used in specifying the models.

Autoregressive models

Idea: predict an output based on the previous outputs.

Let $\lambda = (\lambda_1, \dots, \lambda_T)$ be a time ordered sequence of parameters

An autoregressive Gaussian model for λ is defined by:

a time lag p

a set of coefficients $\{b_1, \dots, b_p\}$

so that

$$\lambda_t = b_1 \lambda_{t-1} + b_2 \lambda_{t-2} + \dots + b_p \lambda_{t-p} + \epsilon_t, \quad \epsilon_t \sim \text{Normal}(0, \sigma_\epsilon^2)$$

equivalently

$$\lambda_t | \lambda_{t-1}, \lambda_{t-2}, \dots, \lambda_{t-p} \sim \text{Normal} \left(\sum_{j=1}^p b_j \lambda_{t-j}, \sigma_\epsilon^2 \right), \quad t = p+1, \dots, T$$

Random walk of order 1, RW(1)

$$\lambda_t = \lambda_{t-1} + \epsilon_t, \quad \lambda_{t-1} = \lambda_{t-2} + \epsilon_{t-1}$$

$$\Rightarrow \lambda_t = \lambda_{t-2} + \epsilon_{t-1} + \epsilon_t$$

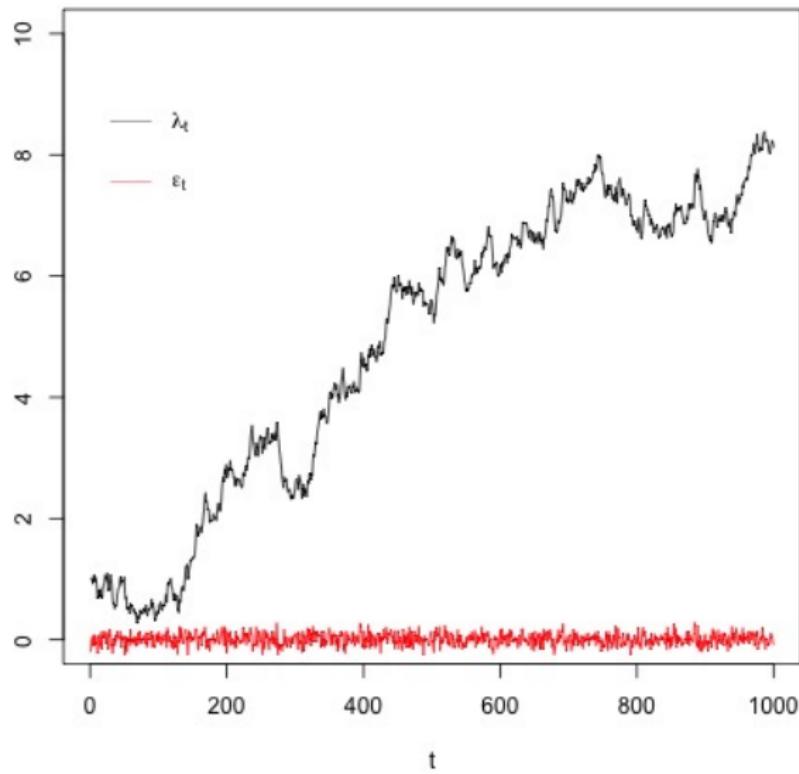
$$\Rightarrow \lambda_t = \lambda_0 + \epsilon_1 + \dots + \epsilon_t$$

Mean $E(\lambda_t) = \lambda_0$

Variance $\text{Var}(\lambda_t) = \text{Var}(\epsilon_1) + \dots + \text{Var}(\epsilon_t) = t\sigma_\epsilon^2 \rightarrow \infty \text{ if } t \rightarrow \infty$

It only models the difference of levels on consecutive time points
 $(\lambda_t - \lambda_{t-1} = \epsilon_t)$

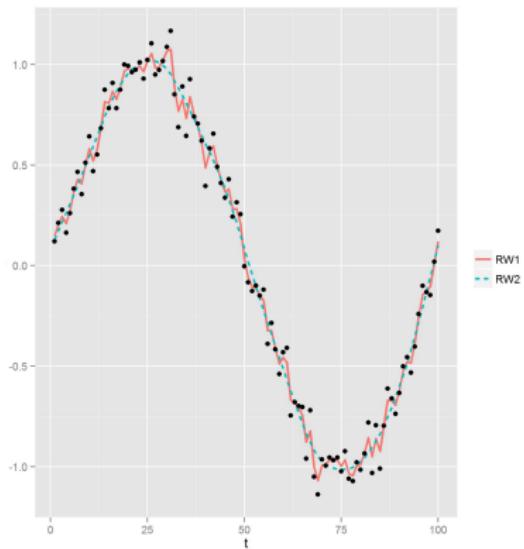
Random walk of order 1, RW(1)



Random walk of order 2, RW(2)

$$\lambda_t = 2\lambda_{t-1} - \lambda_{t-2} + \epsilon_t$$

It only models a linear combination of levels on consecutive time points
 $(\lambda_t - 2\lambda_{t-1} + \lambda_{t-2} = \epsilon_t)$



Simulated data from a sine curve, then RW(1) and RW(2) models fitted

RW(2) model provides greater smoothing than the RW(1) model
(dots=observed data)

Conditional distributions for a RW(1) and RW(2)

The conditional distributions $p(\lambda_t | \boldsymbol{\lambda}_{-t})$, where $\boldsymbol{\lambda}_{-t}$ represent the vector of $\boldsymbol{\lambda}$ with λ_t removed, can be derived.

RW1: The conditional distribution of λ_t involves λ_{t-1} :

$$\lambda_t | \boldsymbol{\lambda}_{-t}, \sigma_\epsilon^2 \sim \text{Normal}(\lambda_{t-1}, \sigma_\epsilon^2)$$

RW2: The conditional distribution of λ_t involves $\lambda_{t-2}, \lambda_{t-1}$:

$$\lambda_t | \boldsymbol{\lambda}_{-t}, \sigma_\epsilon^2 \sim \text{Normal}(2\lambda_{t-1} - \lambda_{t-2}, \sigma_\epsilon^2)$$

Back to disease mapping: Simple additive space-time structure

$$\begin{aligned}y_{it} &\sim \text{Poisson}(\phi_{it} E_{it}) \\ \log \rho_{it} &= \beta_0 + u_i + v_i + \lambda_t + \gamma_t\end{aligned}$$

where

β_0 overall log RR in Georgia over the 11-year period

$v_i \sim \text{Normal}(0, \sigma_v^2)$ spatially unstructured RE

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

$\lambda_t \sim \text{RW}(1)$ temporally structured RE with variance parameter σ_λ^2

$\gamma_t \sim \text{Normal}(0, \sigma_\gamma^2)$ temporally unstructured RE

Adding a space-time interaction

The model can be extended by including **space-time interactions parameters**, δ_{it} .

The interactions can be modelled in different ways depending on which parameters are supposed to interact

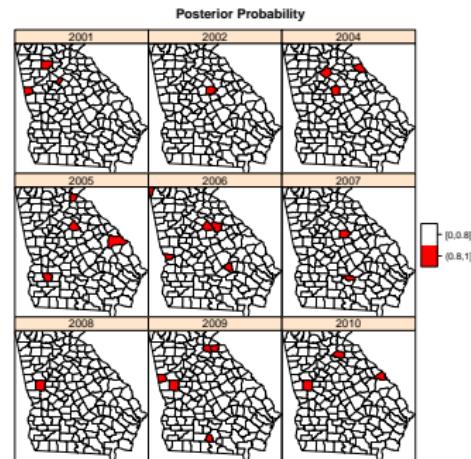
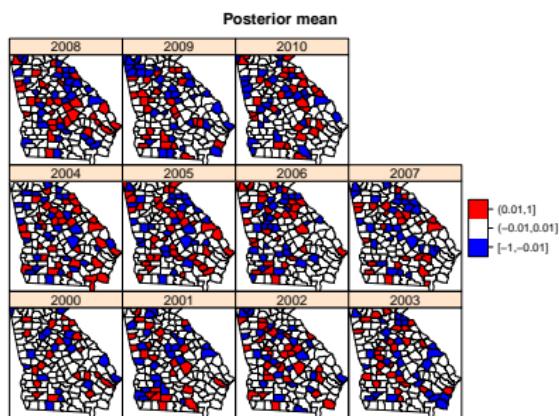
$$\begin{aligned}y_{it} &\sim \text{Poisson}(\phi_{it} E_{it}) \\ \log \phi_{it} &= \beta_0 + v_i + u_i + \lambda_t + \gamma_t + \delta_{it} \\ v_i + u_i &= (\text{BYM} = \text{HET} + \text{ICAR}) \\ \lambda_t &\sim \text{RW}(1) \\ \gamma_t &\sim N(0, \sigma_\gamma^2)\end{aligned}$$

- (i) v_i and γ_t interact: non structured;
- (ii) v_i and λ_t interact: dependence in time but independent for each area;
- (iii) u_i and γ_t interact: dependence in space but independent for each time point;
- (iv) u_i and λ_t interact: completely structured.

Georgia low birth weight - results with space-time interaction

Assuming a type I interaction (non structured)

$$\begin{aligned}y_{it} &\sim \text{Poisson}(\phi_{it} E_{it}) \\ \log \phi_{it} &= \beta_0 + v_i + u_i + \lambda_t + \gamma_t + \delta_{it} \\ &\dots \\ \delta_{it} &\sim N(0, \sigma_\delta^2)\end{aligned}$$



Some more complex models

Lots of work has been done in this area and using INLA as too!
Some models have built up on this approach and have extended it in different directions. Just to name a few:

Goicoa et al. "Age-space-time CAR models in Bayesian disease mapping", Stats in Med. (2015) - extension of a disease mapping in three dimensions applied to prostate cancer mortality in Spain [Goicoa et al., 2016]

Useful shiny app <https://emi-sstcdapp.unavarra.es/>

Gómez-Rubio et al. "Bayesian joint spatio-temporal analysis of multiple diseases", SORT (2019) - extension of a disease mapping for multiple health outcomes, applied to oral, oesophagus and stomach cancer in Spain. They also provide a comparison of MCMC and INLA [Gómez-Rubio et al., 2019]

Summary

Spatial and spatio-temporal models are popular to assess trends and variations at ecological level in epidemiological studies.

We have presented some of the standard structures used in this types of studies which are all available in INLA.

We will now move to the implementation of these models in INLA.

HAVE FUN (and some coffee first!)

References I

-  Besag, J. (1974).
Spatial interaction and the statistical analysis of lattice systems.
Journal of the Royal Statistical Society: Series B (Methodological), 36(2):192–225.
-  Besag, J., York, J., and Mollié, A. (1991).
Bayesian Image Restoration with Two Applications in Spatial Statistics.
The Annals of the Institute of Statistics and Mathematics, 43(1):1–59.
-  Blangiardo, M. and Cameletti, M. (2015).
Spatial and Spatio-temporal Bayesian Models with R-INLA.
Wiley.
-  Clayton, D. and Bernardinelli, L. (1996).
Bayesian methods for mapping disease risk.
In P. Elliott, J. Cuzick, D. E. and Stern, R., editors, *Geographical and Environmental Epidemiology: Methods for Small Area Studies*. Oxford University Press.

References II

-  Gelman, A., Carlin, J., Stern, H., and Rubin, D. (2014).
Bayesian Data Analysis, Third Edition.
Chapman and Hall/CRC.
-  Goicoa, T., Ugarte, M., Etxeberria, J., and Militino, A. (2016).
Age–space–time car models in bayesian disease mapping.
Statistics in medicine, 35(14):2391–2405.
-  Gómez-Rubio, V., Palmí-Perales, F., López-Abente, G., Ramis-Prieto, R., and Fernández-Navarro, P. (2019).
Bayesian joint spatio-temporal analysis of multiple diseases.
SORT-Statistics and Operations Research Transactions, 1(1):51–74.
-  Hoff, P. (2009).
A First Course in Bayesian Statistical Methods.
Springer.

References III

-  Leroux, B. G., Lei, X., and Breslow, N. (2000). Estimation of disease rates in small areas: a new mixed model for spatial dependence. In *Statistical models in epidemiology, the environment, and clinical trials*, pages 179–191. Springer.
-  Martins, T. G., Simpson, D., Lindgren, F., and Rue, H. (2013). Bayesian computing with INLA: New features. *Computational Statistics & Data Analysis*, 67:68–83.
-  Richardson, S., Thomson, A., Best, N., and Elliott, P. (2004). Interpreting posterior relative risk estimates in disease-mapping studies. *Environmental health perspectives*, 112(9):1016–1025.
-  Rue, H. and Held, L. (2005). *Gaussian Markov Random Fields. Theory and Applications*. Chapman & Hall.

References IV



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

Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations.

Journal of the Royal Statistical Society: Series B, 2(71):1–35.



Rue, H., Riebler, A., S., S. H., Illian, J. B., Simpson, D. P., and Lindgren, F. K. (2017).

Bayesian computing with inla: A review.

Annual Review of Statistics and Its Application, 4(1):395–421.