GEOG0125

ADVANCED TOPICS IN SOCIAL AND GEOGRAPHIC DATA SCIENCE

SPATIAL TEMPORAL MODELS & BAYESIAN UPDATING

Dr Anwar Musah (a.musah@ucl.ac.uk)
Lecturer in Social and Geographic Data Science
UCL Geography

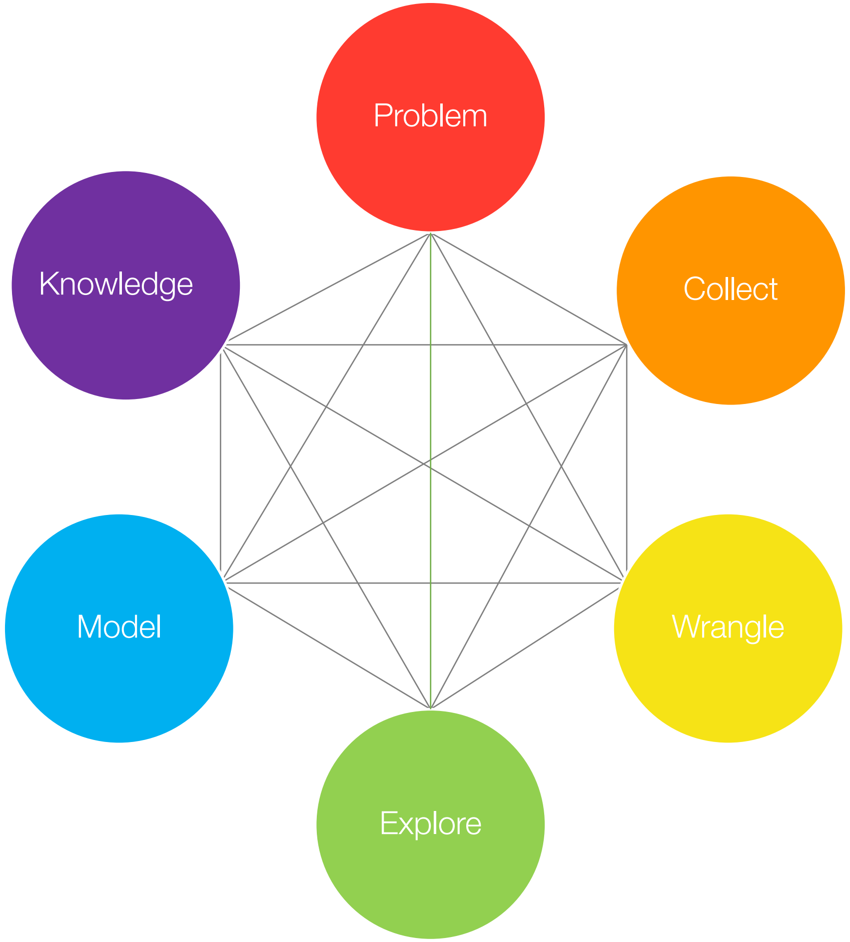
Contents

- A quick recap of Week 8's ICAR model

- Data scenarios
  - ❖ Bayesian Updating for repeated cross-sectional datasets
  - ❖ Spatial Temporal Models for longitudinal datasets

- What is Bayesian Updating:
  - ❖ Let's explore an example: Mosquito infestation in Campina Grande, Brazil
  - ❖ Model framework in Stan

- What are Spatial Temporal Models
  - ❖ Let's explore an example: Fire-related casualties (2010—2019)
  - ❖ Model framework in a different R package called INLA

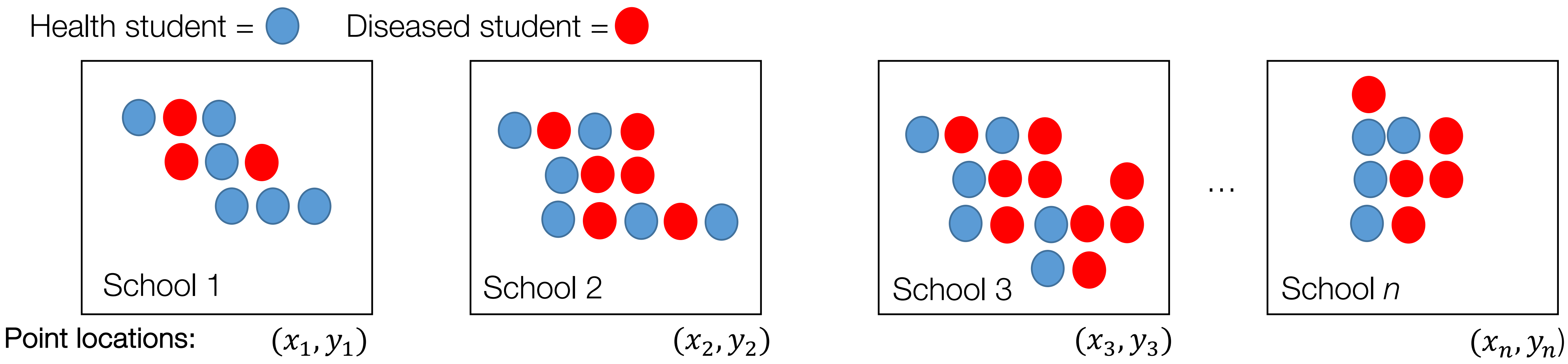


Quick recap on last week's ICARs for spatial risk modelling

We situated hierarchical models within a spatial context:

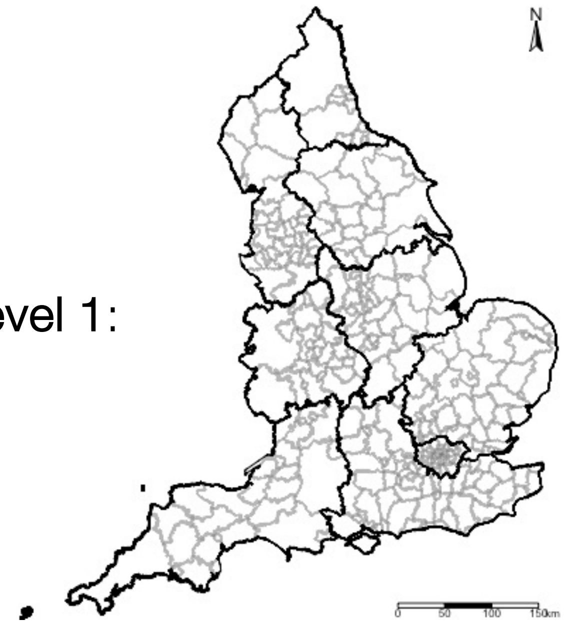
Bayesian inference, are often used in hierarchical modelling, which are models commonly used in the quantification of spatial and spatiotemporal areal data.

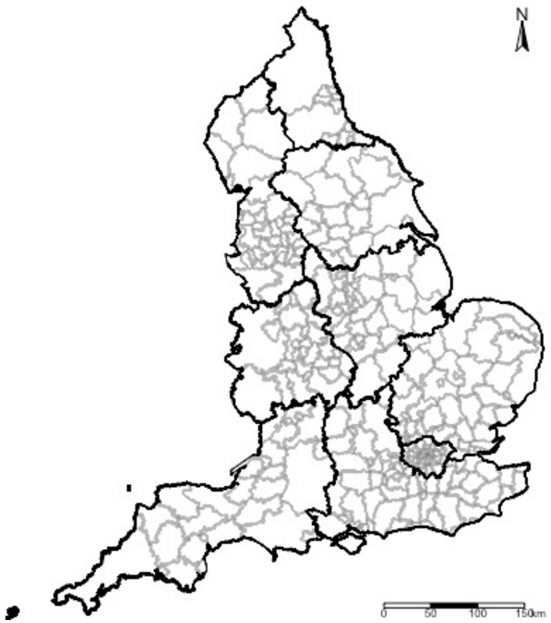
- Bayesian approach are incredibly good with datasets that have a hierarchical structure.
- These are statistical model written in multiple levels (i.e., hierarchical form) to estimate parameters of the posterior distribution
- Example: Intestinal parasitaemia among school children in Tanzania and infection status linked with anaemia

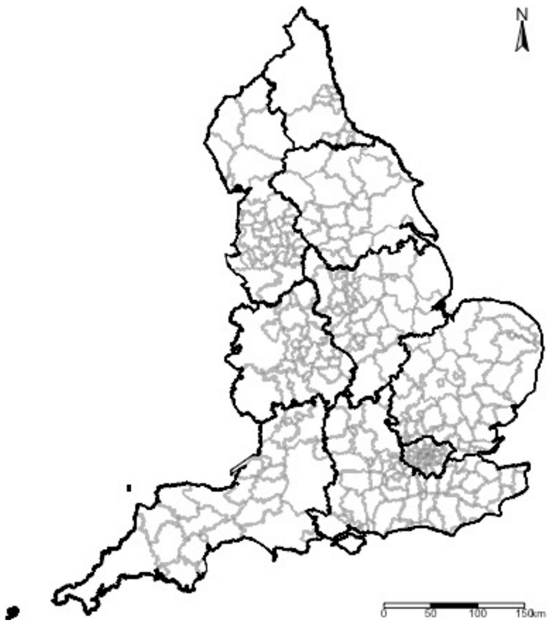


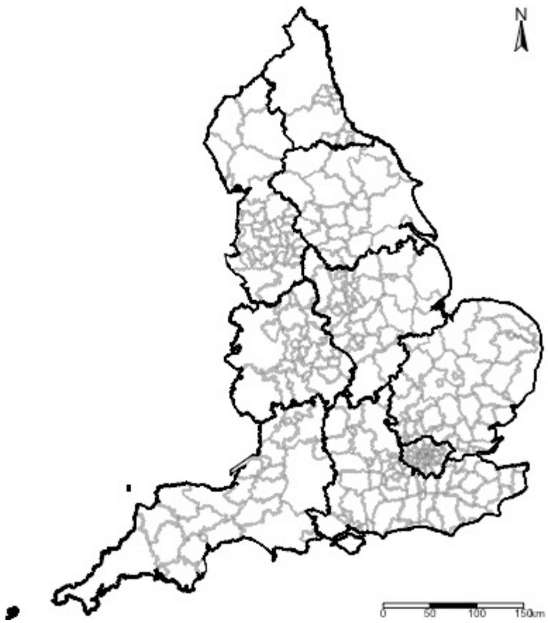
Level 2: $t = 2019$ $t = 2020$ $t = 2021$ $t = 2022$

Level 1:









- These models allow complete flexibility in the estimation of risks - allowing the user to account for space-time interactions

- You can make the model (in contrast to frequentist) to borrow strength across space-time, in order to improve estimation and prediction of an underlying model's feature

Besag-York-Mollie (BYM) (or ICAR models)

This is a popular spatial model which takes into account that the data may potentially be spatially correlated and the observations in the neighbouring areas may be more similar than observations in areas that are distant from each other.

- This is a type of hierarchical model which includes a spatial random effect,
- It is heavily dependent on the neighbourhood adjacency matrix
- ICAR stands for Intrinsic spatial Conditional Auto-Regressive Models
- There are two versions of this model:
  - ❖ <u>BYM model that has a spatial effect term only that's treated a smoothing term (multiplied by an error term)</u>
  - ❖ BYM model that has both a spatial effect term which is treated a structured random effect, and the error term is an unstructured noise
- When fitting data to this type of model – the best choice of the likelihood function (i.e., statistical model) is
  - ❖ <u>Poisson</u> (i.e., aggregated counts (for rates) to areas);
  - ❖ Or otherwise, it's <u>Binomial</u> (aggregate counts (for proportions) to areas).

General model formulation for Spatial ICAR model

Model components

Variables

Y_i are counts of observed cases (outcome)
$X_{i,k}$ independent variables (single variable)
E_i are expected counts of road accident cases (derived from $Y_i R$)
R is the overall rate for the entire study location (not for each area)
r_i is some area-specific rates (this is specified in Poisson statement)

Parameters

α is the overall risk for the entire study area (intercept)
β_k measures the overall associated risk between $X_{i,k}$ and Y_i
ϕ_i are the area-specific spatial random effects
θ_i are the area-specific unstructured random effects
σ an overall error term

Model Calibration

- ρ is the proportion that's set by the user to state the how much variance comes from either ϕ_i or θ_i
- $C_i = \theta_i + \phi_i$ is the combined random effects which is equivalent to $\sigma\left(\sqrt{(1-\rho)}\theta_i + \sqrt{(\rho)}\phi_i\right)$

Notes:
- $\exp(\alpha)$ is the overall risk ratio for study area
- $\exp(\beta)$ is the overall risk ratio for coefficient
- $\exp(\alpha + \sum \beta_k X_{i,k} + C_i\sigma)$ by adding $+C_i\sigma$ to the α allows the risks to vary for each area. By adding $+\sum \beta_k X_{i,k}$ you are also adjusting the estimated risk for the variables.

Full model specification

- Specify likelihood function. The outcome is often counts – thus it will be Poisson (with log as the link function).

$Y_i \sim \text{Poisson}(E_i r_i)$, where $E_i r_i = \lambda_i$

- $\log(\lambda_i) = \alpha + \sum \beta_k X_{i,k} + C_i\sigma + \log(E_i)$
- where $C_i = \theta_i + \phi_i = \left(\sqrt{(1-\rho)}\theta_i + \sqrt{(\rho)}\phi_i\right)$

- Define the priors for the intercept, coefficients and spatial and unstructured random effects as with an ICAR specification

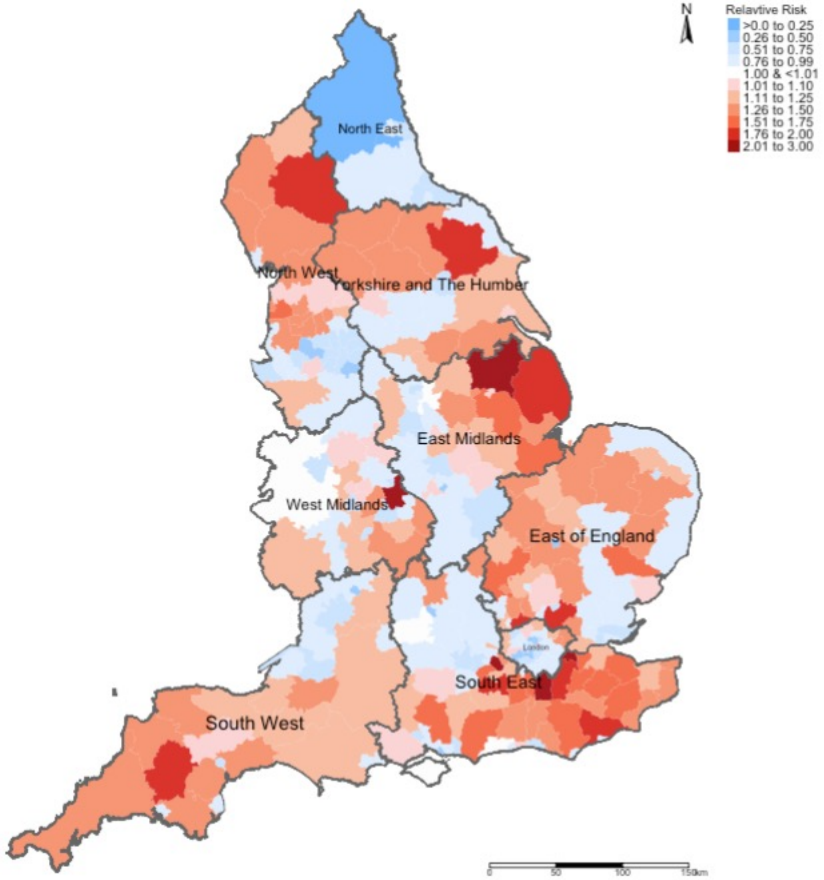
$\alpha \sim \text{norm}(0, 1)$
$\beta \sim \text{norm}(0, 1)$
$\sigma \sim \text{norm}(0, 1)$ (alternatives are $\text{gamma}(0.001, 0.001)$)
$\rho \sim \text{beta}(0.5, 0.5)$
target += -0.5*dot_self(phi[node1] – phi[node2]) (calculates weights)
sum(phi) ~ normal(0, 0.001*N)

- Build Bayesian model

Recall the Bayes' Rule: $P(\theta|Y) \propto P(Y|\theta)P(\theta)$

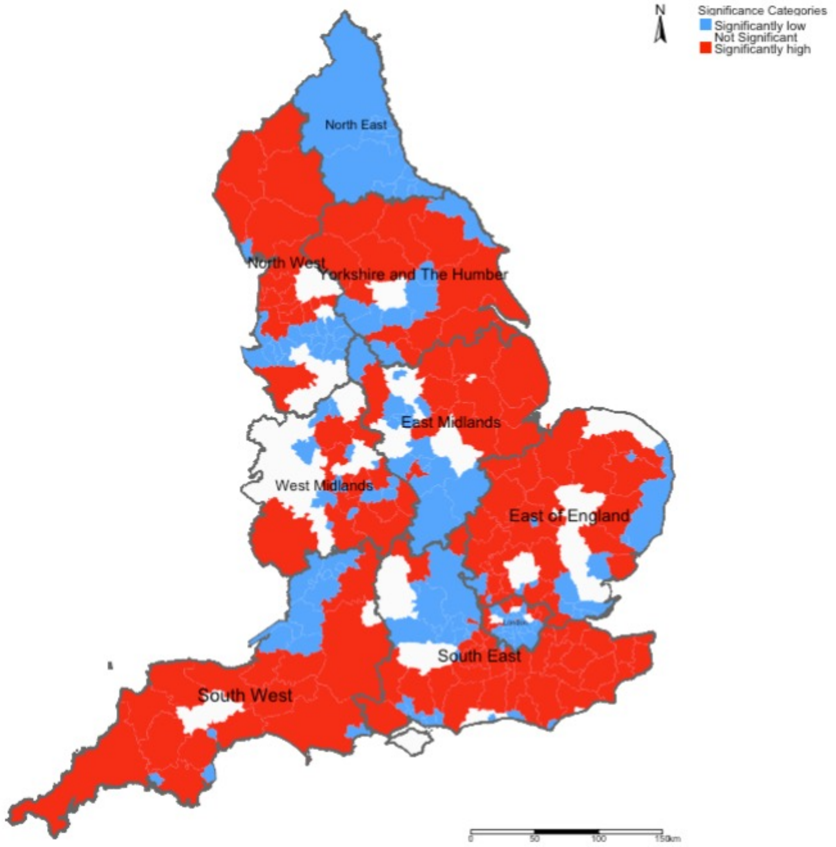
$P(\alpha, \beta_k, \sigma, \varphi_i |\, \lambda_i) \propto P(\,\lambda_i | \alpha, \beta_k, \sigma, \varphi_i)\; P(\alpha)P(\beta_k)P(\sigma)P(\phi_i)P(\rho)$

[1] Relative risk ratios (RR)



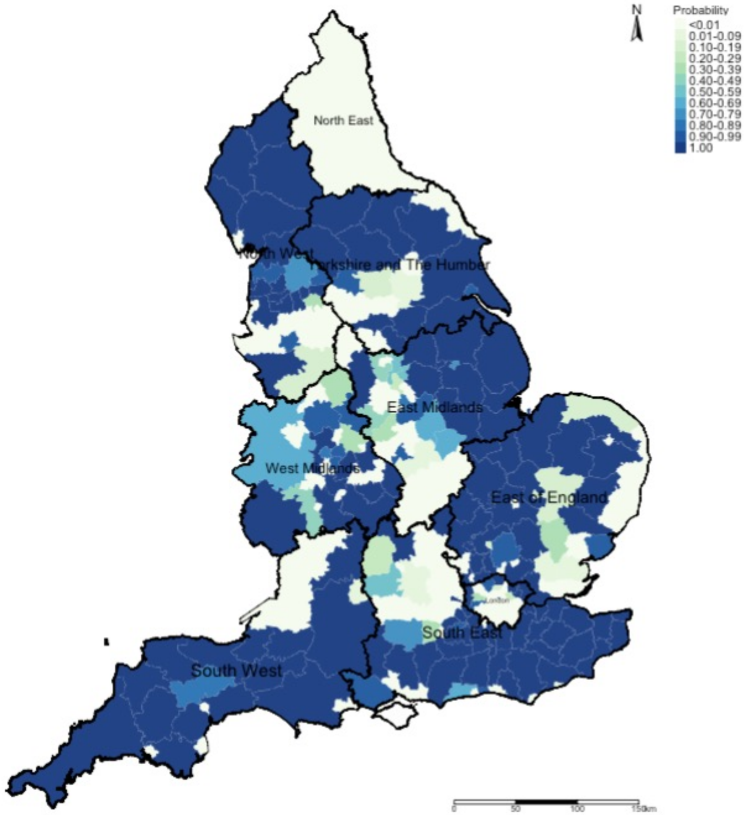
Here, we use this output to describe the burden of an outcome

[2] Statistical Significance



This output is to validate whether the risks described in output [1] are statistically significant or not

[3] Exceedance Probabilities



This output is used to describe the uncertainty that surrounds the risks which we found output [1]

Application a cross-sectional framework

	Survey Periods				
LIRAa	**2013**	**2014**	**2015**	**2016**	**2017**
1	January	January	January	April	January
2	March	March	March	July	$\boxed{\text{April}}$
3	May	May	May	October	July
4	July	July	October		
5	October	October			

Using this snapshot time-period for April to demonstrate the application. This is a Cross-sectional framework

Methodology:

- Population-based ecological study design within cross-sectional (and retrospective) framework

- For covariates, the analysis included:
    1) **WorldClim (4.5km)** (**Maximum temperature** and **Precipitation**) (monthly)
    2) **MOD18A1.061 Terra Vegetation Indices 16-Day Global 500m** to compute neighbourhood levels of vegetation based on the **NDVI** metrics (monthly)
    3) **Worldpop.org (100m)** to extract rasters for urbanisation (which contains binary grids) to compute the fraction of surface that is urbanised for neighbourhoods (yearly).

- **Spatial risk model with Intrinsic Conditional Autoregressive (ICAR) Model**; and to derive new global coefficients for covariates for at that survey period for April 2017, as well as neighbourhood-specific relative risk estimates.

PART 1: Table results that illustrates the GLOBAL association between environmental, climate and anthropogenic factors and risk of infestation in Campina Grande (in LIRAa 2 survey period for April 2017) [Research framework is a cross-sectional study design].

2017	LIRAa 2	
	RR (95% Crl)	Pr(RR>1)
Intercept	1.64 (95% Crl: 0.14 to 7.07)	0.51
Temperature	0.93 (95% Crl: 0.74 to 1.12)	0.23
Precipitation	1.01 (95% Crl: 0.96 to 1.07)	0.73
NDVI	1.09 (95% Crl: 0.71 to 1.60)	0.63
Urbanisation	1.18 (95% Crl: 0.37 to 2.90)	0.52

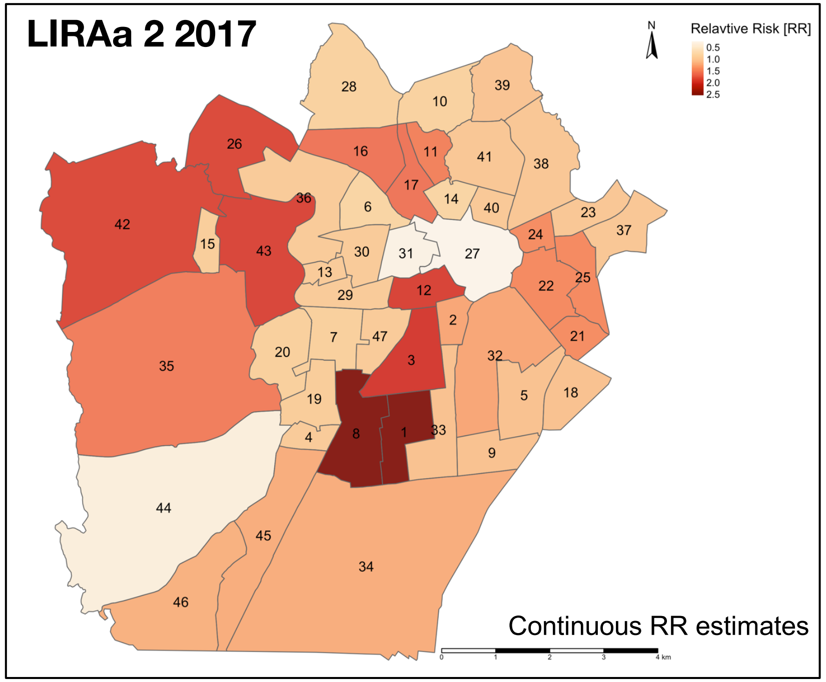
RR: Relative risks; Pr(RR > 1): Exceedance probabilities (the probability that RR being greater than 1)

Interpretation (examples):

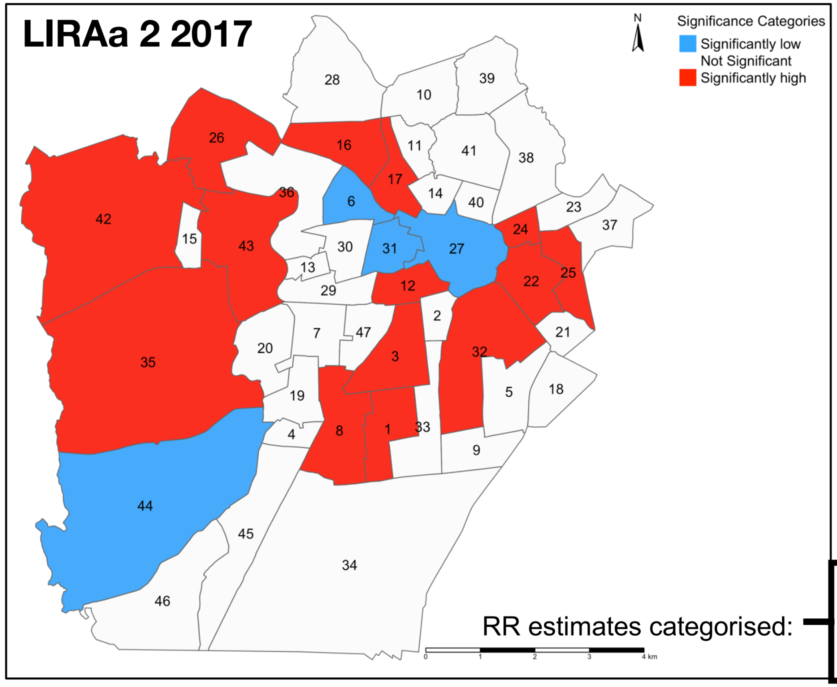
- **Intercept**: The overall risk of mosquito-borne infestation is 1.64 times (or 64%) **higher** in Campina Grande. The overall probability that there's excess risk of infestation (i.e., RR > 1.00) is 51%.
- **Temperature**: In relation to temperature, the risk of mosquito-borne infestation is 0.93 times (or 7%) **lower** in Campina Grande. The probability of observing an excess risk of infestation (i.e., RR > 1.00) in relation to temperature is 23%.
- **Urbanisation index**: In relation to urbanisation, the risk of mosquito-borne infestation is 1.18 times (or 18%) **higher** in Campina Grande. The probability of observing an excess risk of infestation (i.e., RR > 1.00) in relation to urbanisation is 52%.

NOTE: All relative risk estimates have the null value (1) between its lower and upper 95% credibility intervals. While the results, excluding temperature, show an increased risk of infestation – **these are all statistically not significant**.

PART 2: Maps on the left panel illustrates the relative risk (RR) of infestation across neighbourhoods in Campina Grande



PART 3: Maps on the right panel illustrates which neighbourhoods in Campina Grande have RRs that are significantly "low" or "high" risk



RR estimates categorised:

- **RR < 1.00 (Low risk)**
- **RR > 1.00 (High risk)**
- RR = 1.00 (Non-significant risk)

NOTE: There should be a map for exceedance probabilities (this was not shown)

Interpretation:
The following neighbourhoods in Campina Grande numbered 1, 3, 8 and 12 (for example) have RRs that are significantly above 1.00. These are examples of neighbourhoods containing households predicted to be at '**high risk**' of being infested with mosquitoes. Neighbourhoods painted in **RED** need to be monitored for mosquito breeding hotspots to prevent further infestation, which, in turn, can lead to infectious disease outbreaks e.g., Zika or Dengue viruses!

Scenarios

What if:

[1] The dataset are received in streams, or in a ‘repeated’ manner either on a regular or irregular interval

For example, the routine survey for mosquito infestation in Campina Grande, Brazil

	Survey Periods				
LIRAa	**2013**	**2014**	**2015**	**2016**	**2017**
1	January	January	January	April	January
2	March	March	March	July	April
3	May	May	May	October	July
4	July	July	October		
5	October	October			

Note: While the surveys are routine, do you notice how the survey periods are conducted irregular? The study design is **ecological study** design known to be in a **repeated cross-sectional framework**.

When you have a scenario, like this, it would be **improper** to analysis it as **‘true’ longitudinal dataset** or a spatial temporal fashion.

What if:

[2] The dataset are received in a full longitudinal format, where there is no lapses in the time frame i.e., the intervals are ‘regular’

For example, annually report surveillance information on incident fire hazards across Fire Service Areas (FSA) in England, UK

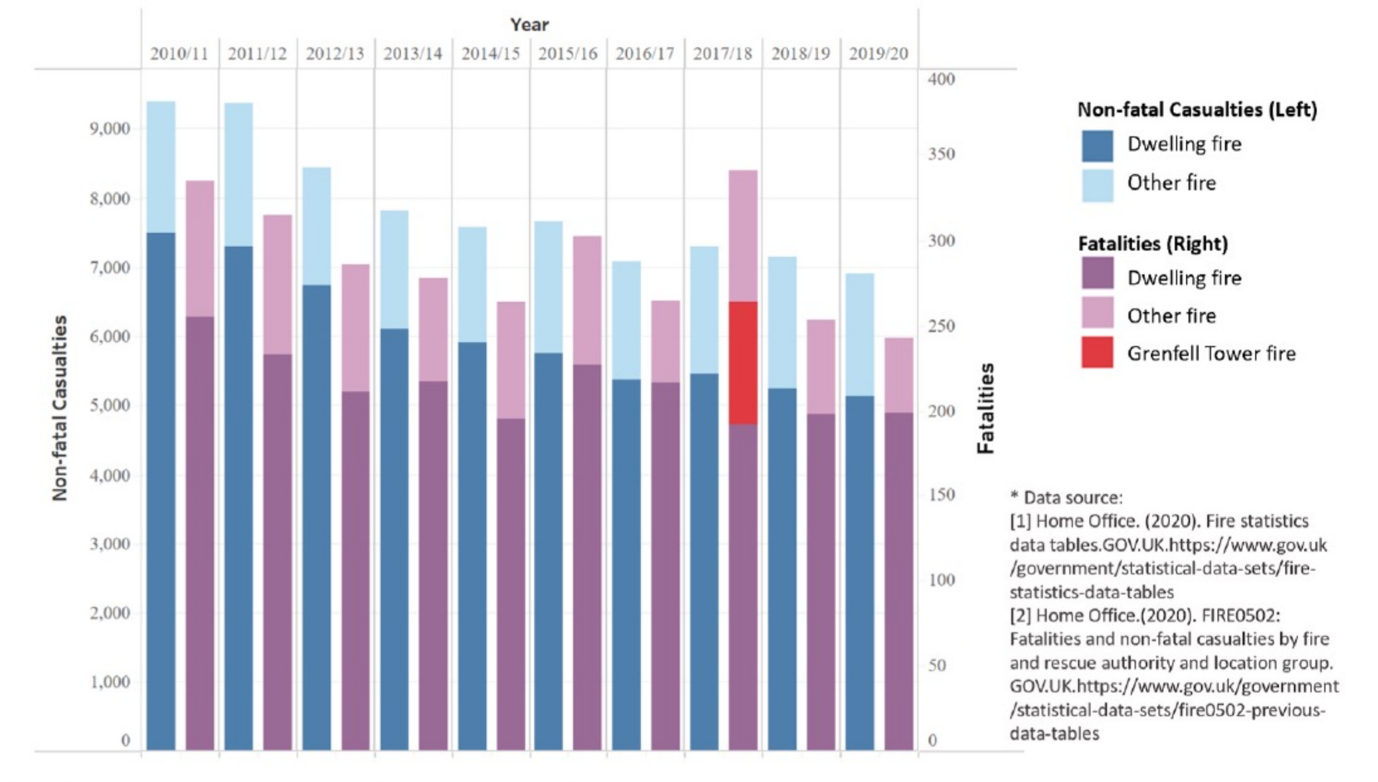


Fig. 1. The bar chart shows the number of fire fatality and non-fatal casualty in England from 2010 to 2019.

Note: The study design is also called an ecological study design known to be in a longitudinal framework.

When you have a scenario, like this, it would be **proper** to analyse it as **‘true’ longitudinal dataset**

Li et al, (2022), Ecological study exploring the geospatial associations between socioeconomic deprivation and fire-related dwelling casualties in England (2010-2019), DOI: https://doi.org/10.1016/j.apgeog.2022.102718

Scenarios

There are two approaches here:

[1]

Bayesian Model Updating

[2]

Spatial Temporal Models

Scenarios

[1]

Bayesian Model Updating

[2]

Spatial Temporal Models

The first scenario is the **best** approach for analysing data that's received in streams on a 'regular' or 'irregular' basis. Suitable for use within a **repeated cross-sectional analytical framework**

Scenarios

[1]

Bayesian Model Updating

[2]

Spatial Temporal Models

The other scenario is the **best** approach for analysing data that's received in longitudinal (which is often retrospective). It has no lapses in its interval. Suitable for use within a **longitudinal analytical framework**

Bayesian Updating

Bayesian Updating

Bayesian updating is a statistical method for revising the estimated probabilities about the events you've observed previously, and this revision is based on new evidence or information (data).

- Bayesian updating technique is excellent choice for survey datasets that are generated routinely in a **repeated cross-sectional** fashion.
- It is used for early warning modelling and off-the-fly predictions as soon as new data is received.
- Bayesian updating needs to be executed in an iterative manner.
- You have a baseline dataset, in which other subsequent data is applied to the Bayesian model for updating. Note that on the initial model with baseline information – you will need to specify the prior distribution for the parameters needed to be estimated. Afterwards, the model recalibrates itself by using the parameter estimates (i.e., posterior distribution) from the previous model run to set them as its new prior.
- The updating will go through a process of a "learning curve" where you will see the estimated parameters evolve in the face of increasing data.

Example: Risk assessment and mapping of infestation in Campina Grande using Bayesian updating

- Campina Grande is into 47 neighbourhoods

- Most recent vector control data (Levantamento Rapid de Indice para Aedes a [LIRAa]): January 2013 to October 2017 (performed 3–5 times in a year)

- Baseline information – the overall number of houses in neighbourhood (as denominators); total number of households detected to be infested with larvae or adult mosquito (i.e., *Aedes aegypti*)

Aims and objectives:

- To quantify the risk trajectories of mosquito infestation on a neighbourhood-level to informs the profile of the neighbourhood (i.e., whether the risks were sustained across the LIRAa periods).

- Determining the set of environmental, climate and anthropogenic risk factors that impact neighbourhood-levels of *Aedes aegypti* infestation in households.

Research Methodology & Study design

	Survey Periods				
LIRAa	**2013**	**2014**	**2015**	**2016**	**2017**
1	January	January	January	April	January
2	March	March	March	July	April
3	May	May	May	October	July
4	July	July	October		
5	October	October			

Using these 6 snapshots demonstrate the application. This is a repeated cross-sectional framework

Methodology:

- Population-based ecological study design within repeated cross-sectional (and retrospective) framework

- For covariates, the analysis included:
  1) **WorldClim (4.5km)** (**Maximum temperature** and **Precipitation**) (monthly)
  2) **MOD18A1.061 Terra Vegetation Indices 16-Day Global 500m** to compute neighbourhood levels of vegetation based on the **NDVI** metrics (monthly)
  3) **Worldpop.org (100m)** to extract rasters for urbanisation (which contains binary grids) to compute the fraction of surface that is urbanised for neighbourhoods (yearly).

- **Spatial risk model with Intrinsic Conditional Autoregressive (ICAR) Model**; and applied **Bayesian updating** to derive new global coefficients for covariates for each LIRA survey, as well as neighbourhood-specific relative risk estimates.

Model formulation for Spatial ICAR modelling and updating [1]

Model components

Variables

Y_i are counts of infected houses in neighbourhoods (outcome)
$X_{i,k}$ independent variables ($k = 4$)
E_i are expected counts of cases infected houses an area
R is the overall rate of infestation in the study area in LIRAa period
r_i is some area-specific rates within that LIRAa period

Parameters

α is the overall risk of infestation for entire study area
β_k measures the overall associated risk between $X_{i,k}$ and Y_i
ϕ_i are the area-specific spatial random effects
θ_i are the area-specific unstructured random effects
σ an overall error term

Model Calibration

- ρ is the proportion that's set by the user to state the how much variance comes from either ϕ_i or θ_i
- $C_i = \theta_i + \phi_i$ is the combined random effects which is equivalent to $\sigma(\sqrt{(1-\rho)}\theta + \sqrt{\rho}\phi)$

Notes:
$\exp(\alpha)$ is the overall risk ratio for study area
$\exp(\beta)$ is the overall risk ratio for coefficient
$\exp(\alpha + \sum \beta_k X_{i,k} + C_i\sigma)$ is risk ratio for each area

Full model specification

- Specify likelihood function. The outcome is often counts – thus it will be Poisson (with log as the link function).

$Y_i \sim \text{Poisson}(E_i r_i)$

- $\log(\lambda_i) = \alpha + \sum \beta_k X_{i,k} + C_i\sigma + \log(E_i)$
- where $C_i = \theta_i + \phi_i = \sigma(\sqrt{(1-\rho)}\theta + \sqrt{\rho}\phi)$

- Define the priors for the intercept, coefficients and spatial and unstructured random effects as with an ICAR specification

$\alpha \sim \text{norm}(0, 1)$
$\beta \sim \text{norm}(0, 1)$
$\sigma \sim \text{norm}(0, 1)$
$\rho \sim \text{beta}(0.5, 0.5)$
$\phi \sim$ ICAR with $\text{norm}(0.001, 0.001 * \text{N})$

- Build Bayesian model

Recall the Bayes' Rule: $P(\theta|Y) \propto P(Y|\theta)P(\theta)$

$$P(\alpha, \beta_k, \sigma, \varphi_i | \lambda_i) \propto P(\lambda_i | \alpha, \beta_k, \sigma, \varphi_i)\, P(\alpha)P(\beta_k)P(\sigma)P(\phi_i)P(\rho)$$

- New models with LIRAa data that follows are updated, this is procedure is known as **Bayesian updating**.

Model formulation for Spatial ICAR modelling and updating [2]

[Step 1] Stan code

```
data {
  int<lower=0> N;
  int<lower=0> N_edges;
  array[N_edges] int<lower=1, upper=N> node1;
  array[N_edges] int<lower=1, upper=N> node2;
  array[N] int<lower=0> Y;
  vector<lower=1>[N] Offset;
  int<lower=1> K;
  matrix[N, K] X;
}

transformed data {
  vector[N] log_Offset = log(Offset);
}

parameters {
  real alpha;                        // intercept
  vector[K] beta;                    // covariates
  real<lower=0> sigma;               // overall standard deviation
  real<lower=0, upper=1> rho;        // proportion unstructured vs. spatially structured variance
  vector[N] theta;                   // unstructured random effects (heterogeneous)
  vector[N] phi;                     // spatial random effects
}

transformed parameters {
  vector[N] combined;
  combined = sqrt(1 - rho) * theta + sqrt(rho) * phi;
}

model {
  Y ~ poisson_log(log_Offset + alpha + X * beta + combined * sigma);

  alpha ~ normal(0.0, 1.0);          // prior for alpha: weakly informative
  beta ~ normal(0.0, 1.0);           // prior for betas: weakly informative
  theta ~ normal(0.0, 1.0);          // prior for theta: weakly informative
  sigma ~ normal(0.0, 1.0);          // prior for sigma: weakly informative
  rho ~ beta(0.5, 0.5);              // prior for rho: pulled for literature

  target += -0.5 * dot_self(phi[node1] - phi[node2]);
  sum(phi) ~ normal(0, 0.001 * N);
}

generated quantities {
  vector[N] eta = alpha + X * beta + combined * sigma;  // compute eta and exponentiate into mu
  vector[N] rr_mu = exp(eta);                            // output the neighbourhood-specific relative risks in mu
  vector[K] rr_beta = exp(beta);                         // output the risk ratios for each coefficient
  real rr_alpha = exp(alpha);                            // output the risk ratios for the intercept
}
```

[Step 2] RStudio code

```
# Create that initial dataset
Lira2016_1 <- infestation_data[infestation_data$Year==2016 & infestation_data$LIRAa==1,]
y=Lira2016_1$InfestedNum
x=Lira2016_1[,c(11,12,13,15)]
e=Lira2016_1$Expected

stan.dataset.baseline <- list(N=n, N_edges=n_edges, node1=nod1, node2=nod2, Y=y, X=x, K=length(x), Offset=e)

Bayesian.model.baseline <- stan("Stan model file.stan",  data=stan.dataset.baseline, iter=60000, chains=6, control = list(max_treedepth = 12), verbose = FALSE, seed = 72182605)
# extract the results for tables and map
```

Note: We use this code to run the model for the first time on the initial dataset to get our baseline results

Model formulation for Spatial ICAR modelling and updating [3]

[Step 3] RStudio code

```
# create the subsequent datasets for the updating
Lira2016_2 <- infestation_data[infestation_data$Year==2016 & infestation_data$LIRAa==2,]
Lira2016_3 <- infestation_data[infestation_data$Year==2016 & infestation_data$LIRAa==3,]
Lira2017_1 <- infestation_data[infestation_data$Year==2017 & infestation_data$LIRAa==1,]
Lira2017_2 <- infestation_data[infestation_data$Year==2017 & infestation_data$LIRAa==2,]
Lira2017_3 <- infestation_data[infestation_data$Year==2017 & infestation_data$LIRAa==3,]

# for Lira2016_2
y=Lira2016_2$InfestedNum
x=Lira2016_2[,c(11,12,13,15)]
e=Lira2016_2$Expected

stan.dataset.update1 <- list(N=n, N_edges=n_edges, node1=nod1, node2=nod2, Y=y, X=x, K=length(x), Offset=e)

# apply updating to next dataset based on posterior estimates from last model
Bayesian.model.update1 = stan(fit=Bayesian.model.baseline,  data=stan.dataset.update1, iter=60000, chains=6, control = list(max_treedepth = 12), verbose = FALSE, seed = 72182605)
# extract the results for tables and map

### REPEAT

# for Lira2016_3
y=Lira2016_3$InfestedNum
x=Lira2016_3[,c(11,12,13,15)]
e=Lira2016_3$Expected

stan.dataset.update2 <- list(N=n, N_edges=n_edges, node1=nod1, node2=nod2, Y=y, X=x, K=length(x), Offset=e)

# again, apply updating to next dataset based on posterior estimates from last model
Bayesian.model.update2 = stan(fit=Bayesian.model.update1,  data=stan.dataset.update2, iter=60000, chains=6, control = list(max_treedepth = 12), verbose = FALSE, seed = 72182605)
# extract the results for tables and map

### RINSE AND REPEAT UNTIL ALL DATASETS ARE COMPLETED
```

Table results illustrates the overall association between environmental, climate and anthropogenic factors and risk of infestation in Campina Grande.

2016	LIRAa 1 (Baseline)		LIRAa 2 (Updated from 2016's LIRAa 1)		LIRAa 3 (Updated from 2016's LIRAa 1 & 2)	
	RR (95% CrI)	Pr(RR>1)	RR (95% CrI)	Pr(RR>1)	RR (95% CrI)	Pr(RR>1)
Intercept	1.53 (95% CrI: 0.13 to 6.54)	0.47	1.55 (95% CrI: 0.13 to 6.71)	0.47	1.45 (95% CrI: 0.12 to 6.26)	0.45
Temperature	0.94 (95% CrI: 0.82 to 1.07)	0.18	0.97 (95% CrI: 0.87 to 1.09)	0.33	0.95 (95% CrI: 0.86 to 1.05)	0.17
Precipitation	1.02 (95% CrI: 0.98 to 1.07)	0.87	1.02 (95% CrI: 0.92 to 1.12)	0.67	1.31 (95% CrI: 0.65 to 2.39)	0.75
NDVI	1.01 (95% CrI: 0.87 to 1.16)	0.52	1.02 (95% CrI: 0.78 to 1.31)	0.54	1.02 (95% CrI: 0.59 to 1.62)	0.48
Urbanisation	1.12 (95% CrI: 0.64 to 1.84)	0.62	1.19 (95% CrI: 0.64 to 2.02)	0.68	1.55 (95% CrI: 0.81 to 2.69)	0.91

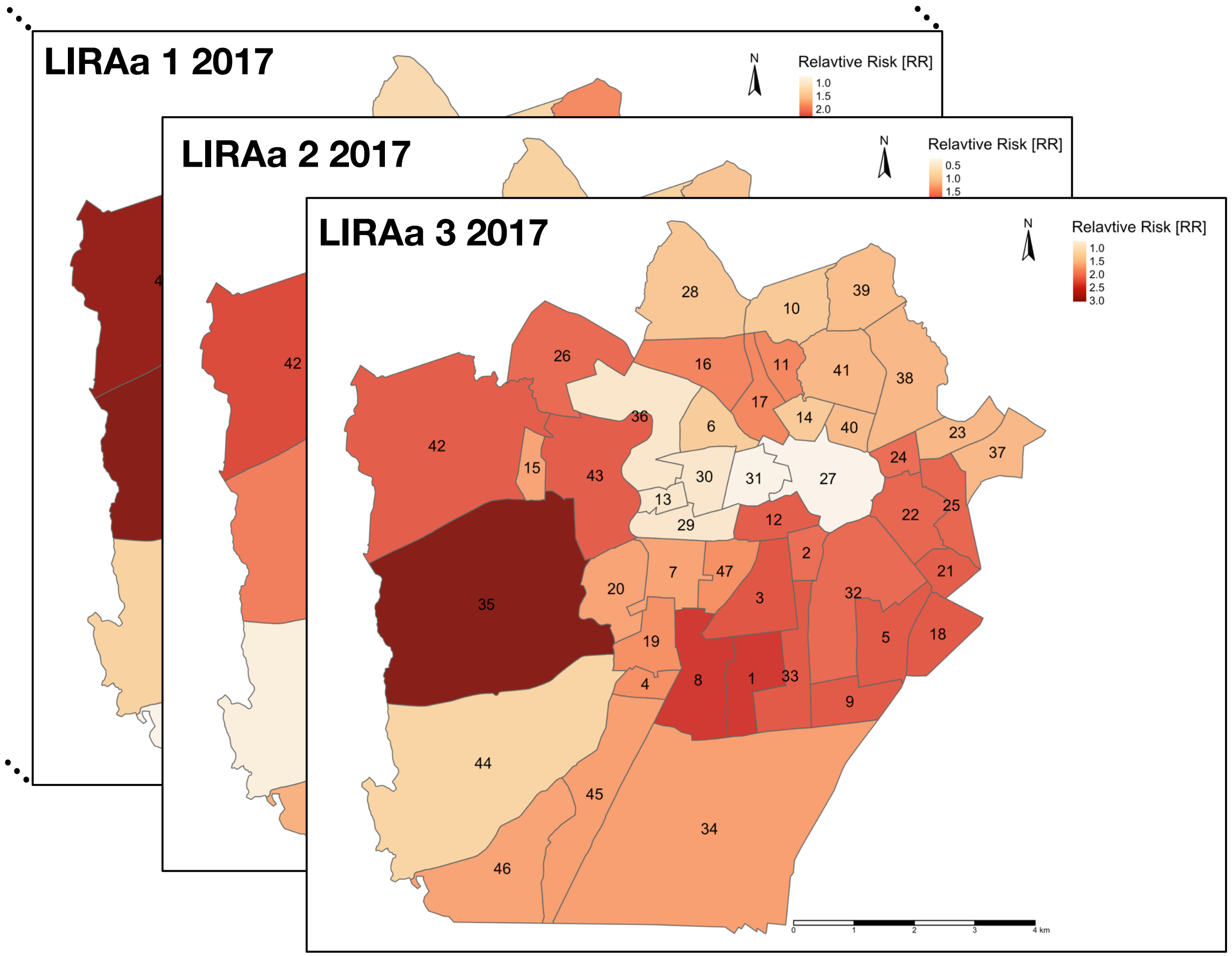
2017	LIRAa 1 (Updated from 2016's LIRAa 1, 2 & 3)		LIRAa 2 (Updated from 2016's LIRAa 1, 2 & 3, and 2017's LIRAa 1)		LIRAa 3 (Updated from 2016's LIRAa 1, 2 & 3, and 2017's LIRAa 1 and 2)	
	RR (95% CrI)	Pr(RR>1)	RR (95% CrI)	Pr(RR>1)	RR (95% CrI)	Pr(RR>1)
Intercept	1.47 (95% CrI: 0.13 to 6.27)	0.45	1.64 (95% CrI: 0.14 to 7.07)	0.51	1.81 (95% CrI: 0.15 to 7.81)	0.53
Temperature	0.92 (95% CrI: 0.82 to 1.03)	0.09	0.93 (95% CrI: 0.74 to 1.12)	0.23	1.01 (95% CrI: 0.88 to 1.15)	0.57
Precipitation	1.15 (95% CrI: 1.03 to 1.28)	0.99	1.01 (95% CrI: 0.96 to 1.07)	0.73	1.00 (95% CrI: 0.98 to 1.01)	0.61
NDVI	0.93 (95% CrI: 0.62 to 1.33)	0.32	1.09 (95% CrI: 0.71 to 1.60)	0.63	0.94 (95% CrI: 0.84 to 1.06)	0.16
Urbanisation	1.19 (95% CrI: 0.57 to 2.20)	0.64	1.18 (95% CrI: 0.37 to 2.90)	0.52	0.82 (95% CrI: 0.45 to 1.37)	0.19

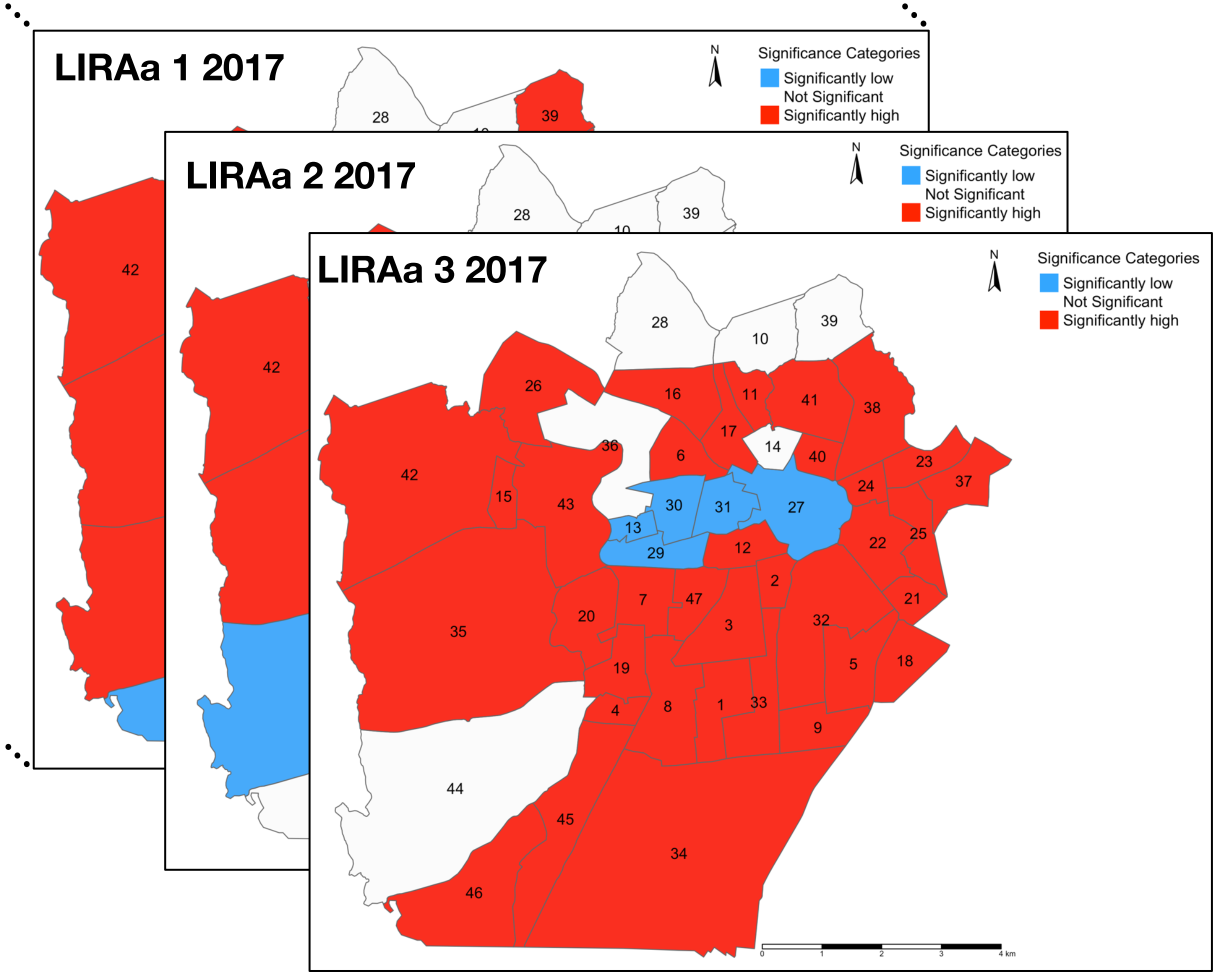
RR: Relative risks

Pr(RR > 1): Exceedance probabilities (the probability that RR being greater than 1)

Maps on the left panel illustrates the relative risk (RR) of infestation across neighbourhoods in Campina Grande

Maps on the right panel illustrates which neighbourhoods in Campina Grande have RRs that are significantly “low” or “high” risk



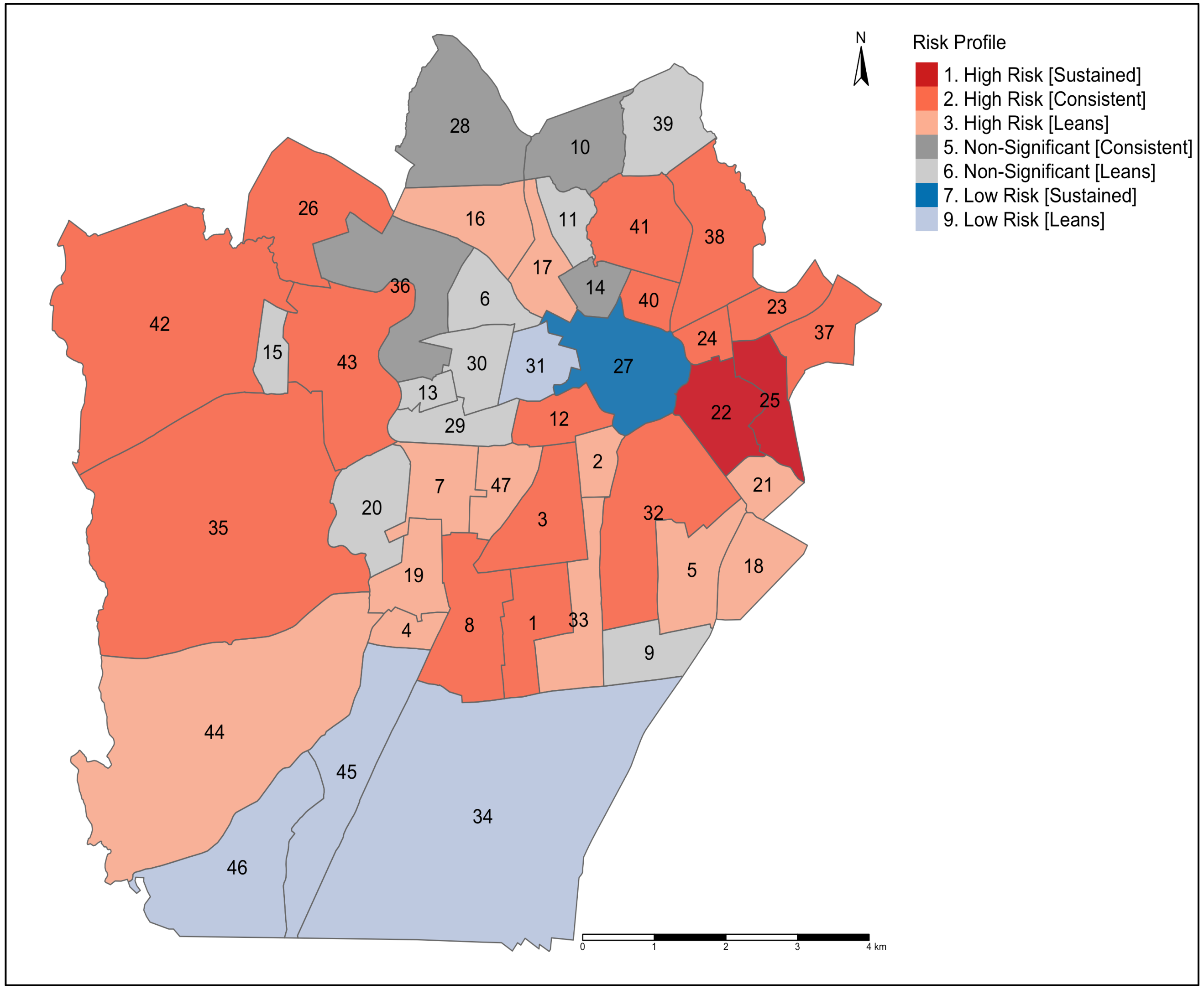


RR < 1.00 (Low risk)
RR > 1.00 (High risk)
RR = 1.00 (Non-significant risk)

Modelling the risk trajectories and charting them across these LIRAa periods

Profiling the neighbourhoods accordingly to examine where these significant risks of infestation are sustained

ID	Neighbourhood	LIRAa 1 04/16	LIRAa 2 07/16	LIRAa 3 10/16	LIRAa 1 01/17	LIRAa 2 04/17	LIRAa 3 07/17	Profile
1	Jardim Paulistano							2. High Risk [Consistent]
2	Estação Velha							3. High Risk [Leans]
3	Liberdade							2. High Risk [Consistent]
4	Presidente Médici							3. High Risk [Leans]
5	Sandra Cavalcante							3. High Risk [Leans]
6	Monte Santo							6. Non-Significant [Leans]
7	Santa Rosa							3. High Risk [Leans]
8	Cruzeiro							2. High Risk [Consistent]
9	Itararé							6. Non-Significant [Leans]
10	Jardim Continental							5. Non-Significant [Consistent]
11	Louzeiro							6. Non-Significant [Leans]
12	São José							2. High Risk [Consistent]
13	Pedregal							6. Non-Significant [Leans]
14	Conceição							5. Non-Significant [Consistent]
15	Ramadinha							6. Non-Significant [Leans]
16	Jeremias							3. High Risk [Leans]
17	Palmeira							3. High Risk [Leans]
18	Vila Cabral							3. High Risk [Leans]
19	Santa Cruz							3. High Risk [Leans]
20	Dinamérica							6. Non-Significant [Leans]
21	Mirante							3. High Risk [Leans]
22	José Pinheiro							1. High Risk [Sustained]
23	Castelo Branco							2. High Risk [Consistent]
24	Santo Antônio							2. High Risk [Consistent]
25	Monte Castelo							1. High Risk [Sustained]
26	Novo Bodocongó							2. High Risk [Consistent]
27	Centro							7. Low Risk [Sustained]
28	Cuités							5. Non-Significant [Consistent]
29	Centenário							6. Non-Significant [Leans]
30	Bela Vista							6. Non-Significant [Leans]
31	Prata							9. Low Risk [Leans]
32	Catolé							2. High Risk [Consistent]
33	Tambor							3. High Risk [Leans]
34	Velame							9. Low Risk [Leans]
35	Malvinas							2. High Risk [Consistent]
36	Universitário							5. Non-Significant [Consistent]
37	Nova Brasília							2. High Risk [Consistent]
38	Jardim Tavares							2. High Risk [Consistent]
39	Nações							6. Non-Significant [Leans]
40	Lauritzen							2. High Risk [Consistent]
41	Alto Branco							2. High Risk [Consistent]
42	Serrotão							2. High Risk [Consistent]
43	Bodocongó							2. High Risk [Consistent]
44	Três Irmãs							3. High Risk [Leans]
45	Acácio Figueiredo							9. Low Risk [Leans]
46	Cidades							9. Low Risk [Leans]
47	Quarenta							3. High Risk [Leans]



I will show you how to pull this off in the practicals!

Spatial Temporal Models

Software for Bayesian methods beyond Stan

- INLA (Integrated Nested Laplace Approximation)

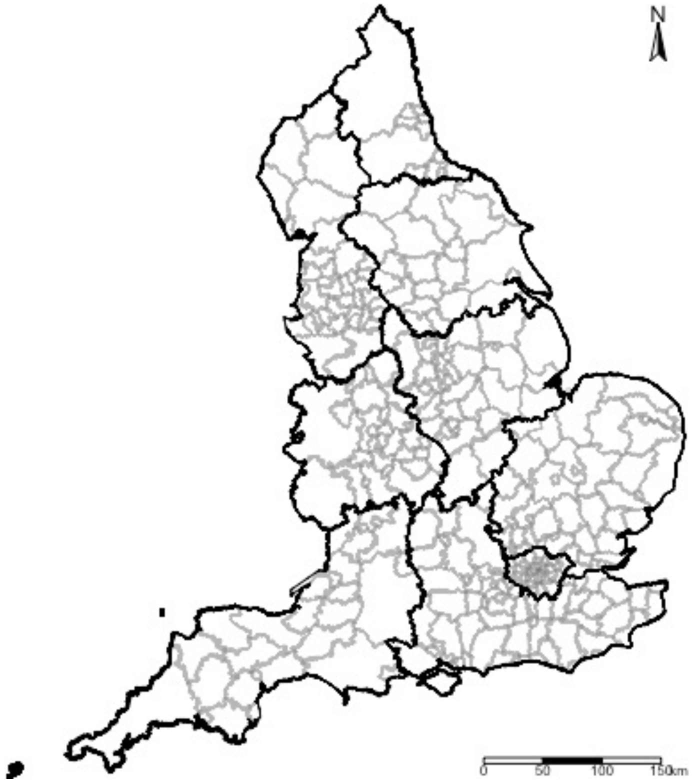
INLA allows to a perform approximate Bayesian inference in latent Gaussian models, ranging from a tonne of generalized linear (mixed) models to various spatial models (e.g., Besag-York-Mollie & SPDEs)

- Thanks to R-INLA Project (http://www.r-inla.org) – it has made learning Bayesian modelling very easy and accessible to anyone with basic knowledge of statistics

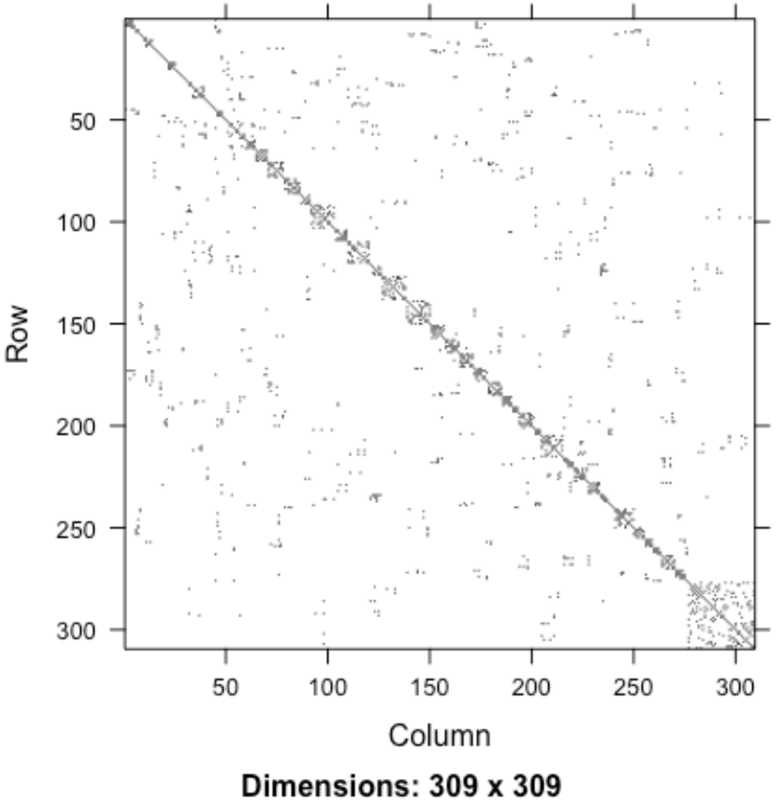
- The codes are streamed-lined and infinitely easier to code than in Stan

- It's a lot efficient when it comes to speed

309 areal units (i) followed over (j) times



Adjacency matrix for Local Authority area u_i



We will use this to construct our spatial random effects unlike Stan, INLA uses an adjacency matrix

Model formulation

$$Y_{ij} \sim \text{Poisson}(E_{ij}, \theta_{ij})$$

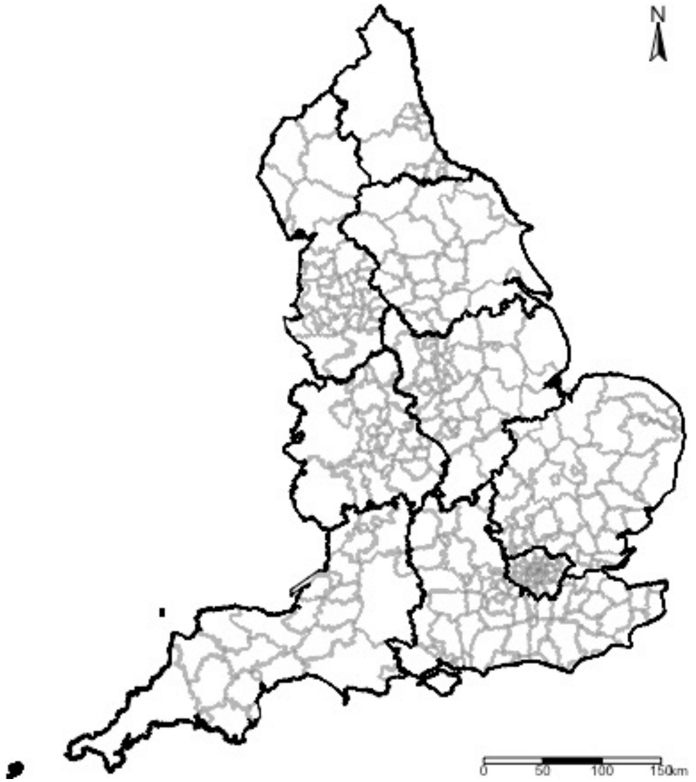
$$\log(\theta_{ij}) = \alpha + u_i + v_i + (X_{k,ij}\beta_{k,ij})t_j + \log(E_{ij})$$

<<<

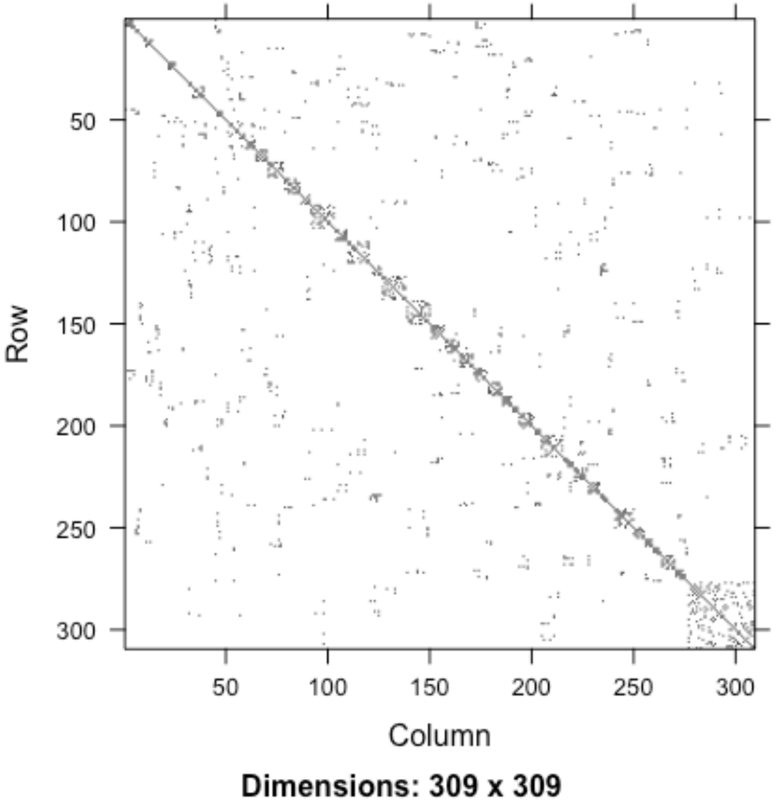
Model components

Y_{ij} is observed counts of cases in area i at a particular time j
E_{ij} is expected number of cases in area i at a particular time j
α is the overall risk of the study area
$(u_i + v_i)$ structured and unstructured spatial random effects
t_j random effect caused by time j
$X_{k,ij}$ independent variables
$\beta_{k,ij}$ coefficient for our independent variables

309 areal units (i) followed over (j) times



Adjacency matrix for Local Authority area u_i



We will use this to construct our spatial random effects unlike Stan, INLA uses an adjacency matrix

Model formulation

$$Y_{ij} \sim \text{Poisson}(E_{ij}, \theta_{ij})$$

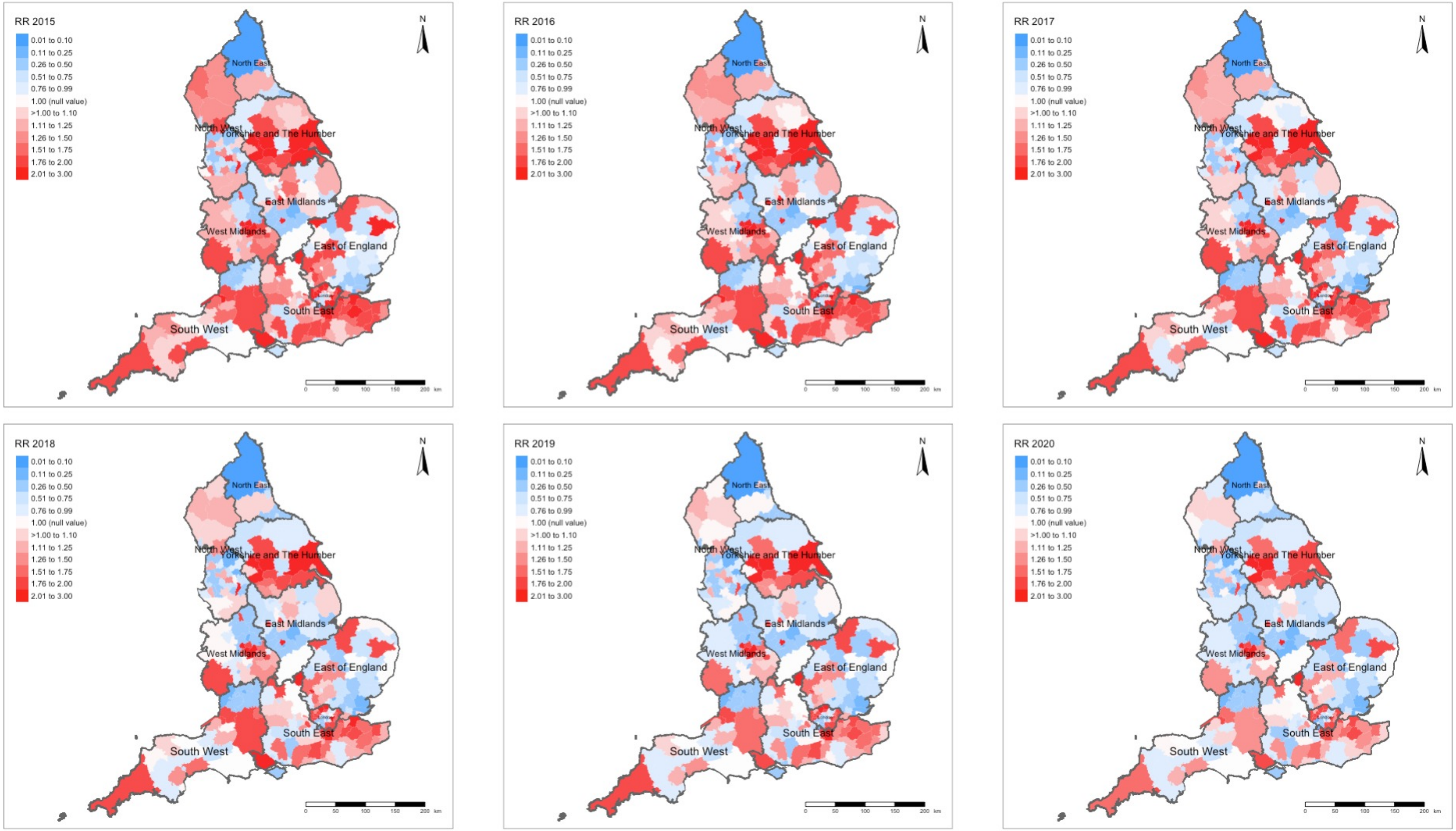
$$\log(\theta_{ij}) = \alpha + u_i + v_i + (X_{k,ij}\beta_{k,ij})t_j + \log(E_{ij})$$

```
formula = Y ~ 1 + f(AreaID, model = "bym2", graph = g, hyper = prior) +
        f(AreaID1, timeID, model = "bym2") + timeID
```

The code is quite simple

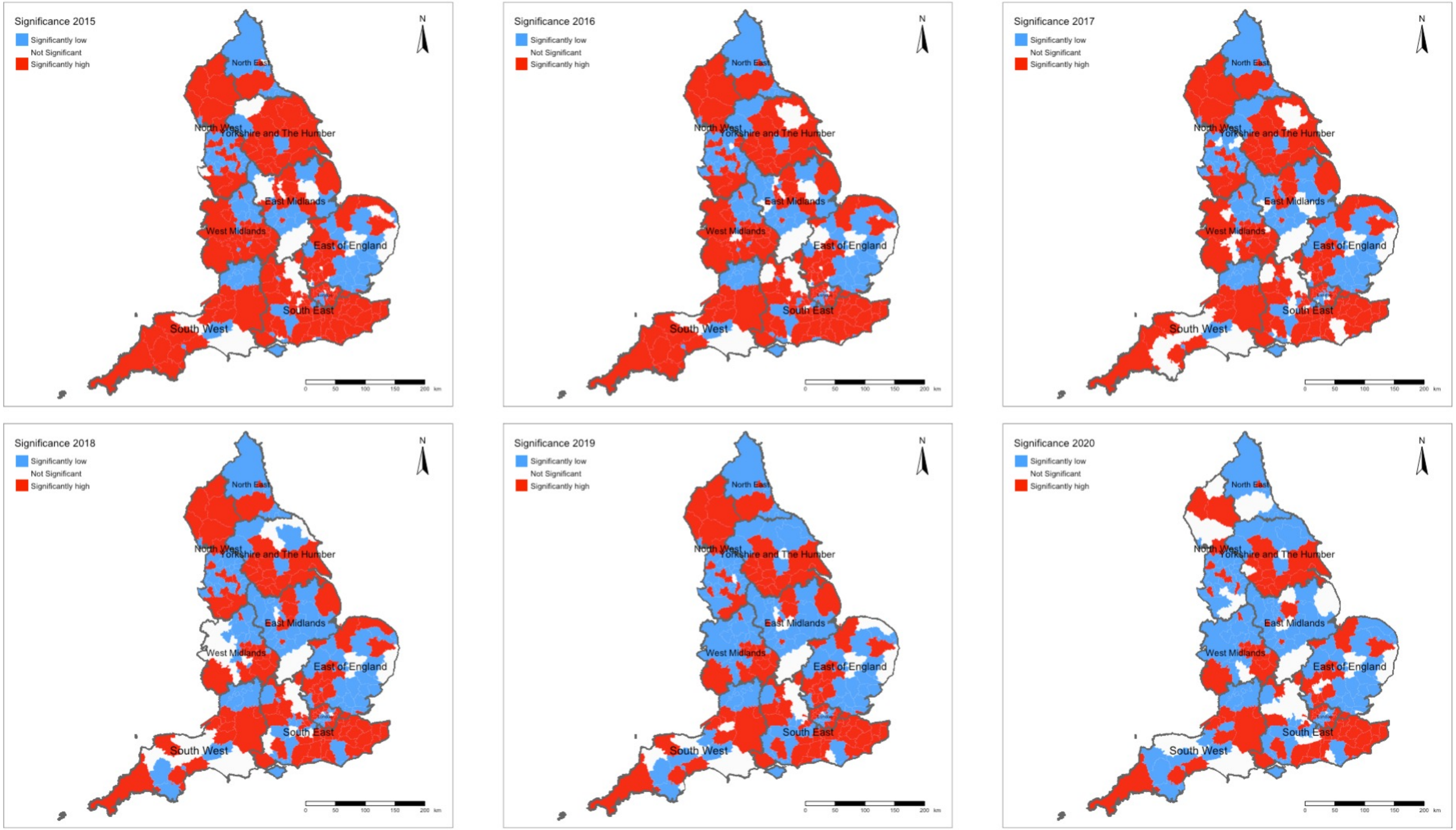
We can get our relative risks, significance, and exceedance probabilities for this code

Example: Spatiotemporal Risks of Road-related casualties in England 2015-2020 [1]

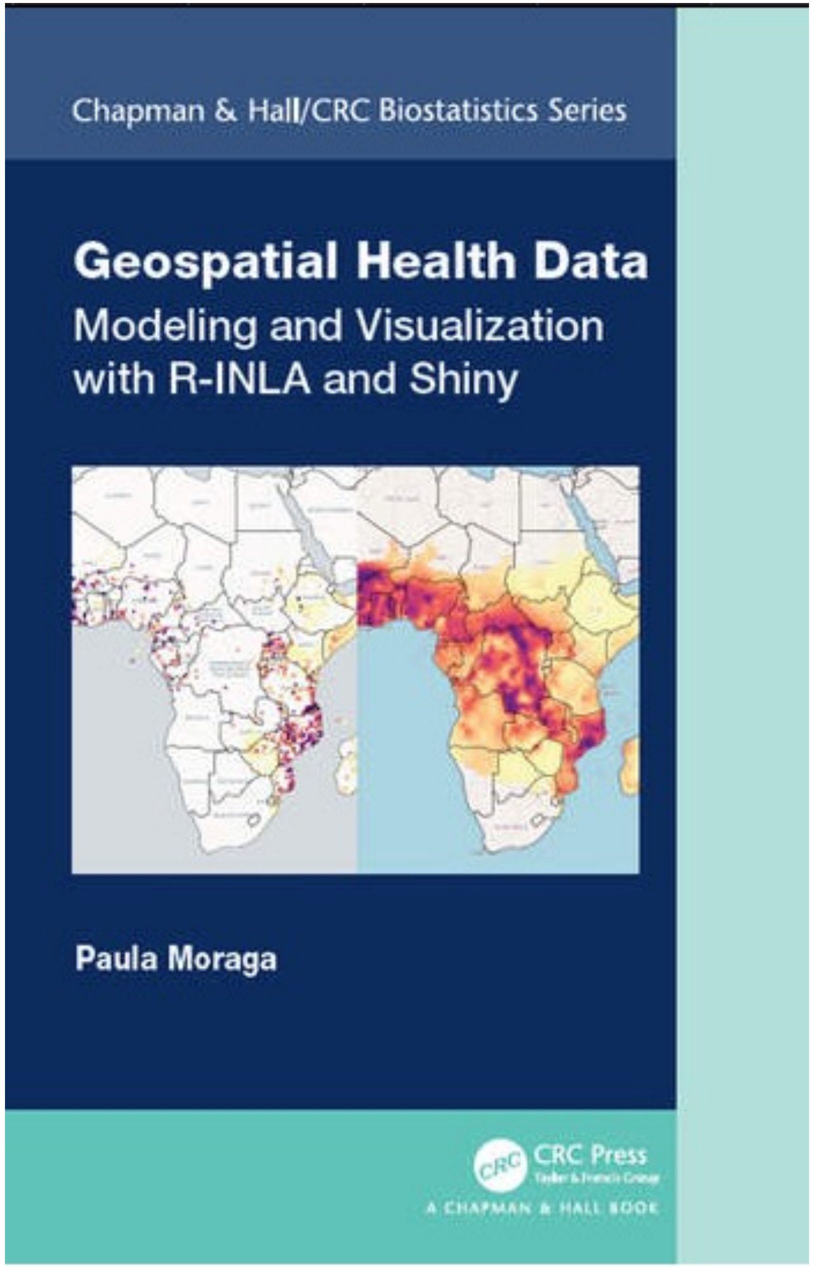


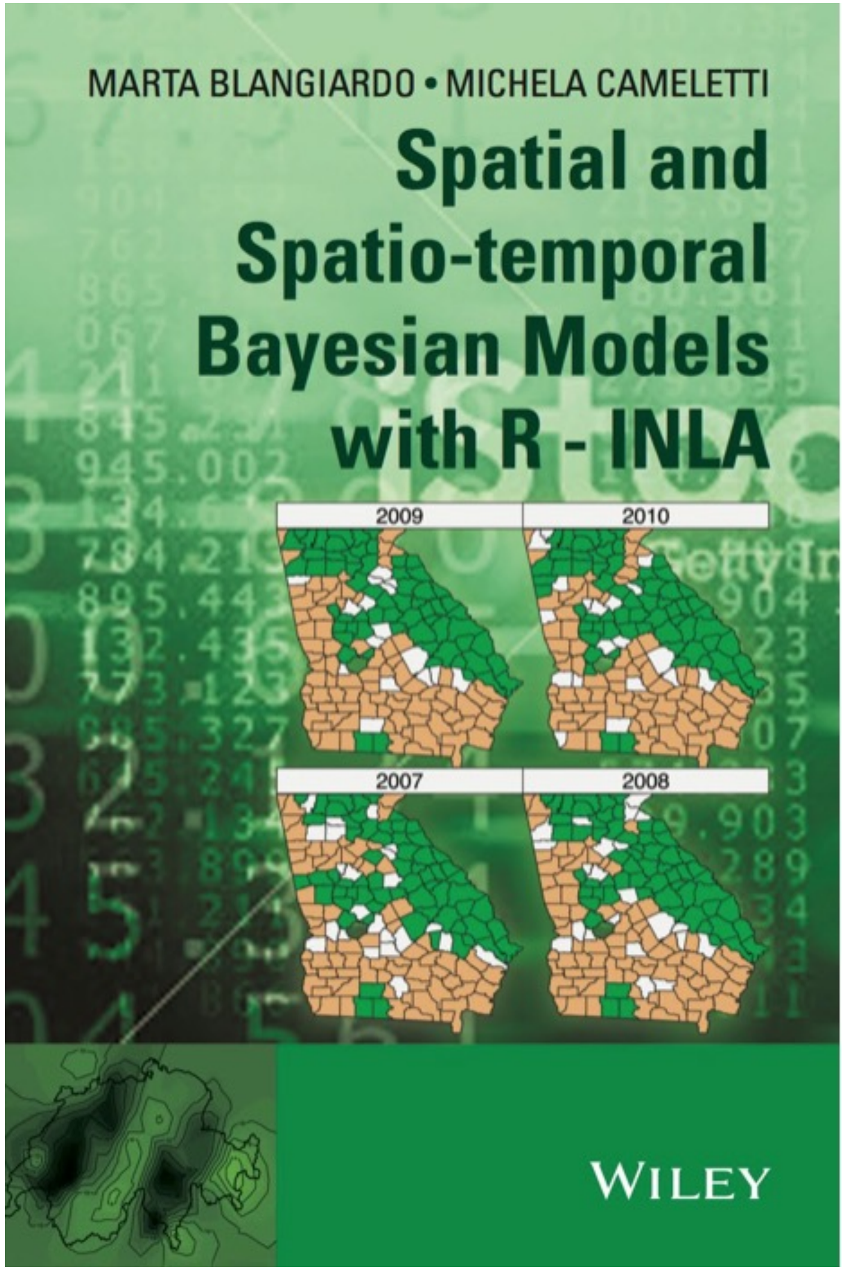
Spatially and temporally varying relative risk estimates

Example: Spatiotemporal Risks of Road-related casualties in England 2015-2020 [2]



Spatially and temporally varying significance categories


Chapman & Hall/CRC Biostatistics Series
Geospatial Health Data
Modeling and Visualization with R-INLA and Shiny
Paula Moraga
CRC Press
A CHAPMAN & HALL BOOK


MARTA BLANGIARDO • MICHELA CAMELETTI
Spatial and Spatio-temporal Bayesian Models with R - INLA
2009
2010
2007
2008
WILEY

Example: Fire-related casualties in England (2010-19) with spatial temporal risk modelling

Table results illustrates the overall association between socioeconomic deprivation factors and risk of fire-related casualties in England (2010-19).

Multivariate spatiotemporal Bayesian regression models that explores the overall association with deprivation indexes with dwelling fire-related SCRs and random effects.

Domains of Deprivation in England	Adjusted Relative Risk (95% Credibility Intervals)		
	Estimates	Percentage	
Living Environment	1.241 (1.164–1.329)	+24.1%	(16.4%–32.9%)
Education, Skills & Training Deprivation	1.181 (1.124–1.245)	+18.1%	(12.4%–24.5%)
Housing & Barriers to Public Services	1.137 (1.094–1.184)	+13.7%	(9.4%–18.4%)
Crime	1.010 (1.007–1.013)	+1.0%	(0.7%–1.3%)
Random effects	**Median (95% Credibility Intervals)**		
τ^2: estimate of temporally varying spatial variation	0.035 (0.028–0.044)		
ρ_S: estimate of spatial autocorrelation	0.014 (0.001 to o.065)		
ρ_T: estimate of temporal autocorrelation	0.936 (0.885–0.986)		

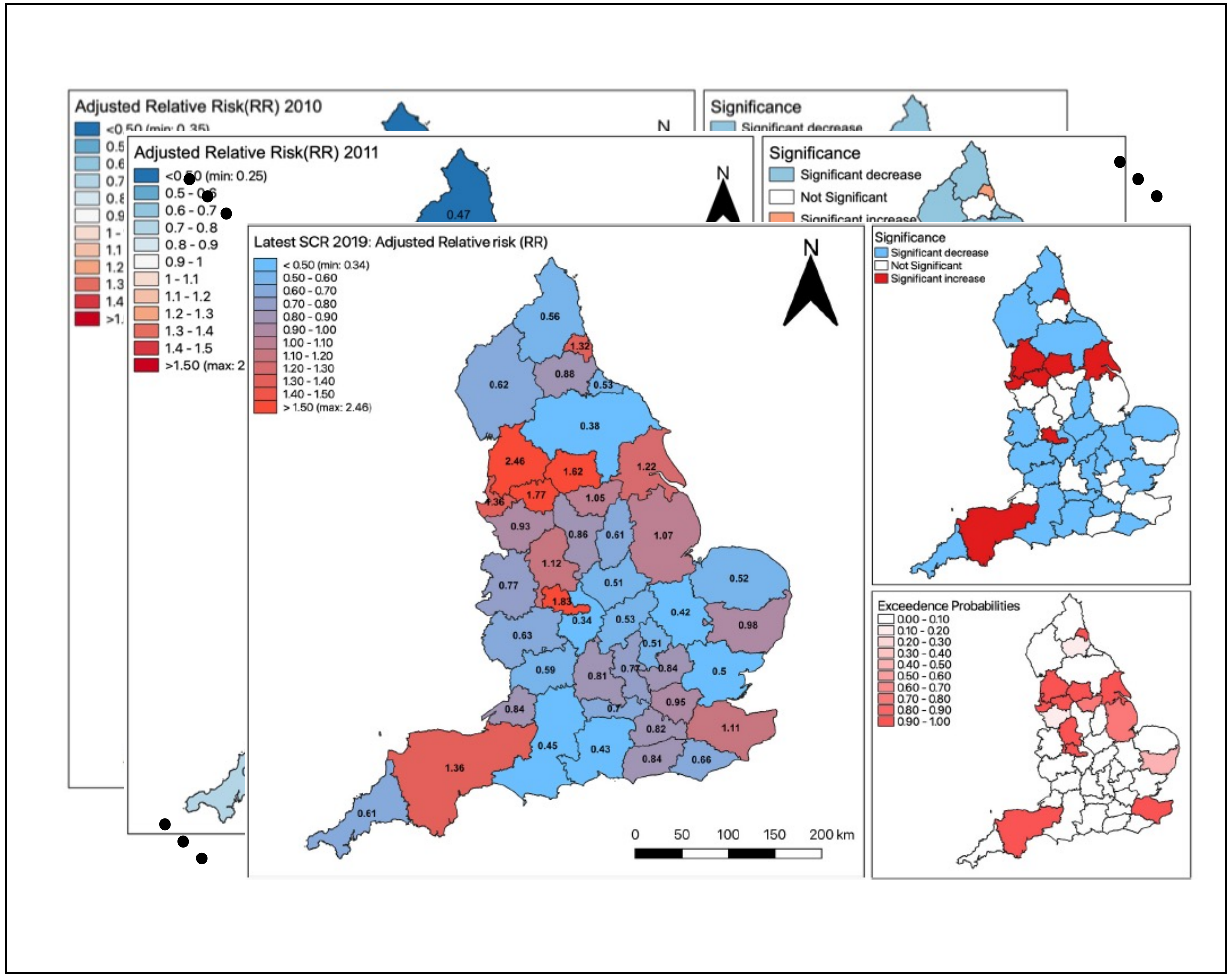
Source: Li et al, (2022), Ecological study exploring the geospatial associations between socioeconomic deprivation and fire-related dwelling casualties in England (2010-2019),
DOI: https://doi.org/10.1016/j.apgeog.2022.102718

Interpretation:

- **Living environment**: In the context of this socioeconomic deprivation domain – **increased levels significantly increases the risk of fire-related casualties by 24.1% (1.241 times higher).**
- **Barriers to housing and public service**: In the context of this socioeconomic deprivation domain – **increased levels significantly increases the risk of fire-related casualties by 13.7% (1.137 times higher).**
- **Education, Skills & Training**: In the context of this socioeconomic deprivation domain – **increased levels significantly increases the risk of fire-related casualties by 18.1% (1.181 times higher).**
- **Crime**: In the context of this socioeconomic deprivation domain – **increased levels marginally increases the risk of fire-related casualties by 1.0% (1.01 times higher).** While this is statistically significant, the significance can be seen as something that is marginal.

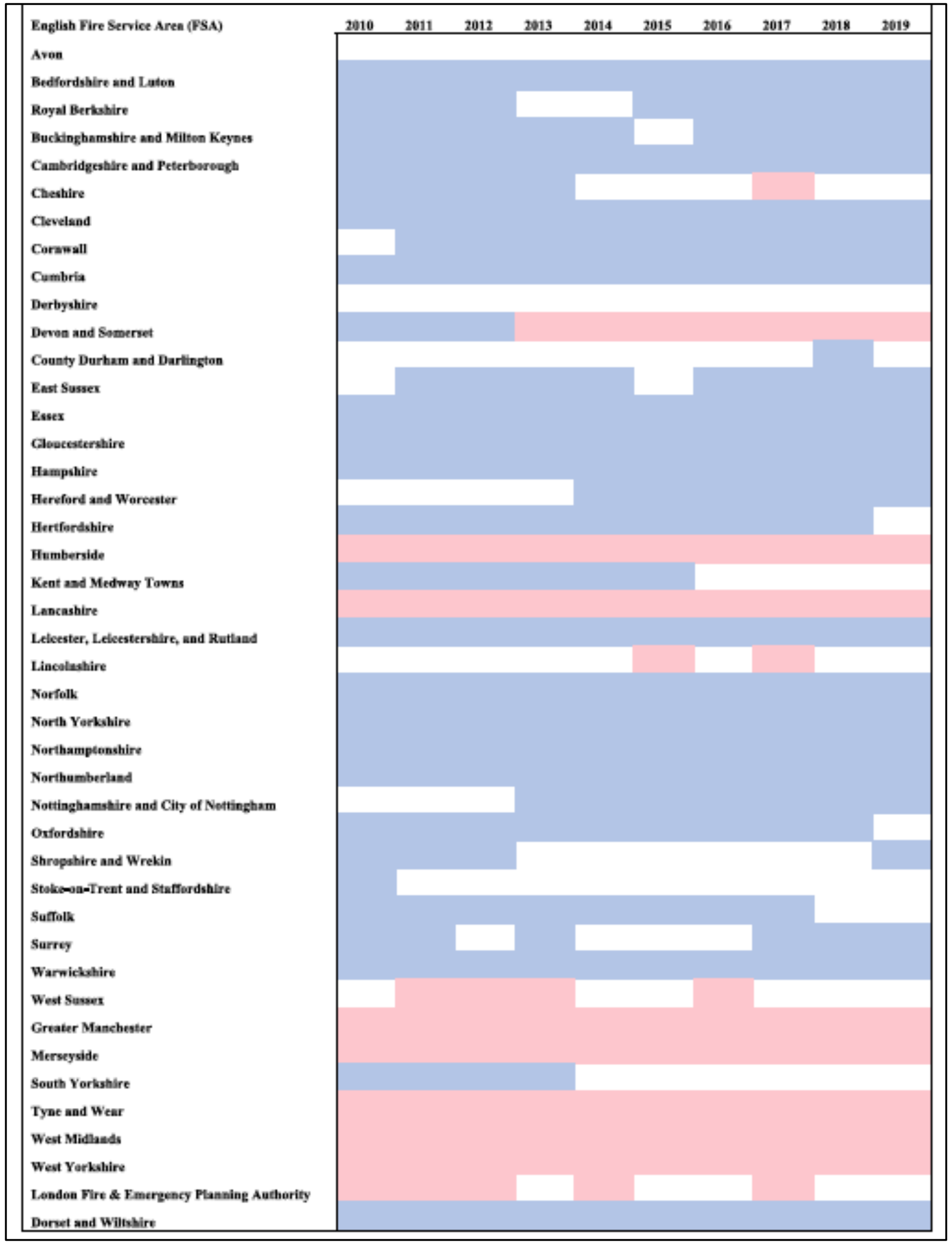
NOTE: The random effects are reported in this instance to show the variation in risk. Most of it is captured in the time component (0.936 = 93.6%) meaning that there is a very strong temporal autocorrelation in the data.

Spatial and temporal variation in risk of fire-related casualties



>>>

Translate maps to show significant risk trajectories



Interpretation: On a year-on-year basis, the increased risks are significantly sustained throughout the 10-year period for Humberside, Lancaster, Greater Manchester, Merseyside, Tyne and Wear, West Midlands, West Yorkshire, and Devon & Somerset.

Source: Li et al, (2022), Ecological study exploring the geospatial associations between socioeconomic deprivation and fire-related dwelling casualties in England (2010-2019). DOI: https://doi.org/10.1016/j.apgeog.2022.102718

Any questions?