Bayesian Spatial Conditional Overdispersion Models

Mabel Morales Otero

Supervisor: Prof. Vicente Núñez-Antón
University of the Basque Country (UPV), Bilbao

Department of Quantitative Methods, Faculty of Economics and Business

mabel.morales@ehu.eus

Overview

1. Background
   - Overdispersion
   - Overdispersion models
   - Spatial count data
2. Spatial conditional overdispersion models
   - Spatial conditional overdispersion models
   - Generalized spatial conditional overdispersion models
3. Application
   - Colombian infant mortality dataset
     - Fitted models and results
     - Comparison to the Besag–York–Mollié (BYM) models
   - COVID-19 incidence in Flanders dataset
     - Fitted models and results
4. Current work and future extensions

Overdispersion

Generalized Linear Models (GLM) assume that the response variables Y_i, for $i = 1, \ldots, n$, follow a distribution which can be normal, Poisson, binomial or any other

Poisson regression model

Let $Y_i \sim \text{Poi}(\mu_i)$, $i = 1, \ldots, n$, then we assume that μ_i follows the regression model:

$$\log(\mu_i) = \mathbf{X}_i' \boldsymbol{\beta},$$

where $\mathbf{X}_i$ is the $k \times 1$ vector of explanatory variables for the i-th observation and $\boldsymbol{\beta}$ is the $k \times 1$ vector of unknown regression parameters

$\Longrightarrow$ Property known as equidispersion: $\text{E}(Y_i) = \text{Var}(Y_i) = \mu_i$

$\Longrightarrow$ **Overdispersion** would occur when $\text{Var}(Y_i) > \mu_i$

Causes of overdispersion

Among the main causes of overdispersion we have:

- Omission of relevant variable or term in the model
- Nonconstant variance of the response variable
- When the independence assumption does not hold $\Longrightarrow$ Very common in spatial data

$\Longrightarrow$ Standard errors may be underestimated and the inferential processes may be incorrect

$\Longrightarrow$ This issue must be taken into account in order to obtain reliable inference processes for the estimated parameters in the proposed model

Overdispersion models for Poisson responses

Normal Poisson model

Given a random effect $\nu_i \sim N(0, \tau)$, we assume that $(Y_i|\nu_i) \sim \text{Poi}(\mu_i)$, where the mean follows the regression model:

$$\log(\mu_i) = \mathbf{X}_i' \boldsymbol{\beta} + \nu_i$$

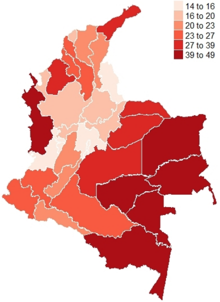
The variance can be approximated so that $Var(Y_i) \approx \mu_i + \tau\mu_i^2$. Since the dispersion parameter is positive, that is, $\tau > 0$, then $Var(Y_i) > \mu_i \Longrightarrow$ Larger than the one specified by the Poisson model $\Longrightarrow$ Overdispersion

Negative binomial (NB2) model

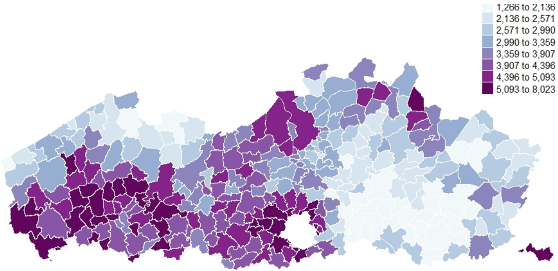
Assume that $Y_i \sim \text{NB}(\tau/(\tau + \mu_i), \tau),\ i = 1, \ldots, n$, with mean $\text{E}(Y_i) = \mu_i$ following the regression model: $\log(\mu_i) = \mathbf{X}_i' \boldsymbol{\beta}$

The variance, $\text{Var}(Y_i) = \mu_i + \tau^{-1}\mu_i^2$ is larger than the one that would have been specified with the Poisson model, since the dispersion parameter is positive, $\tau > 0$

Spatial count data



(a) Infant mortality in Colombia for the year 2005, by department



(b) COVID-19 incidence in Flanders' municipalities from September 2020 to January 2021

$\Longrightarrow$ **Spatial autocorrelation**: observations in locations that are closer in space tend to show similar values

$\Longrightarrow$ Could be one of the causes of overdispersion

$\Longrightarrow$ Must also be taken into account

Spatial conditional overdispersion models

Spatial conditional overdispersion models (Cepeda-Cuervo et al., 2018)

Let Y_i, for $i = 1, \ldots, n$ represent counts for n regions $\Longrightarrow$The spatial conditional overdispersion models assume that $(Y_i|Y_{\sim i})$ follows a conditional overdispersed distribution: $f(y_i|y_{\sim i})$, for $i = 1, \ldots, n$

$Y_{\sim i}$ is defined as the set of values in all of the neighbors of the i-th region, except for the i-th region itself

The conditional mean follows a given regression structure that includes covariates and the spatial lag of the response variable, with a spatial parameter that allows to account for the intensity of the spatial dependence that may be present in the data

Spatial neighborhood structure

- Normally specified for a set of n regions, by the $n \times n$ spatial weights matrix $\mathbf{W} = [w_{ij}]$, where the w_{ij}'s are the weights that represent the strength of the dependence between regions i and j
- $\mathbf{W} = [w_{ij}]$ is usually standardized by rows $\Longrightarrow$ $w^*_{ij} = \dfrac{w_{ij}}{\sum_{j=1}^{n} w_{ij}}$, so that $\sum_{j=1}^{n} w^*_{ij} = 1$
- For region i, $\mathbf{W}_i\mathbf{y}$ is the spatial lag of the response variable $\mathbf{Y}$, defined as the product of the $1 \times n$ vector corresponding to the ith row of the weights matrix $\mathbf{W}$, $\mathbf{W}_i$, and the $n \times 1$ vector of observations, $\mathbf{y}$ $\Longrightarrow$ $\mathbf{W}_i\mathbf{y} = \sum_{j=1}^{n} w^*_{ij} y_j$ is a weighted average of the values at neighboring regions
- There are two main ways of specifying the w_{ij}'s: Adjacency (based on boundaries) and distance

Spatial neighborhood structure: Contiguity

Specifications for $\mathbf{W} = [w_{ij}]$ based on adjacency or contiguity:

- Contiguity of order 1: $w_{ij} = 1$ if region j is adjacent or a neighbor to region i, and 0 otherwise

  $\Longrightarrow$ One way of defining adjacency is by assuming that regions i and j are be neighbors if they share at least one point in their boundaries

- Contiguity of order 2: $w_{ij} = 1$ if regions i and j share a common neighbor, and 0 otherwise $\Longrightarrow$ Higher order can be specified and can be cumulative if we include lower order neighbors.

Spatial neighborhood structure: Distance

Let s_i be the center point (or centroid) of region i, with coordinates $(x_i, y_i) \Longrightarrow$ We could define the distance between regions i and j, for example, as the Euclidean distance between their centroids:

$$d_{ij} = \|s_i - s_j\| = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}$$

Specifications for $\mathbf{W} = [w_{ij}]$ based on distance:

- Inverse distance: $w_{ij} = d_{ij}^{-1}$
- Negative exponential: $w_{ij} = \exp(-d_{ij})$
- Distance band: $w_{ij} = 1$ if $d_{ij} < h$ for a chosen threshold h, and 0 otherwise

Spatial conditional overdispersion models for Poisson responses

Spatial conditional normal Poisson model

Assume that $(Y_i|Y_{\sim i}, \nu_i) \sim \text{Poi}(\mu_i)$, with conditional mean $\mathrm{E}(Y_i|Y_{\sim i}, \nu_i) = \mu_i$ following the regression model:

$$\log(\mu_i) = \mathbf{X}_i'\boldsymbol{\beta} + \rho\mathbf{W}_i\mathbf{y} + \nu_i, \quad \text{where} \quad \nu_i \sim N(0, \tau)$$

Spatial conditional negative binomial model

Assume that $(Y_i|Y_{\sim i}) \sim \text{NB}(\tau/(\tau + \mu_i), \tau)$, with conditional mean $\mathrm{E}(Y_i|Y_{\sim i}) = \mu_i$ following the regression model:

$$\log(\mu_i) = \mathbf{X}_i'\boldsymbol{\beta} + \rho\mathbf{W}_i\mathbf{y}$$

$\Longrightarrow$ These models assume a constant overdispersion, and there are cases where the dispersion in the data can vary among groups or observations

Generalized spatial conditional overdispersion models (Cepeda-Cuervo et al., 2018)

$(Y_i|Y_{\sim i})$, $i = 1, \ldots, n$, follows a conditional overdispersed distribution denoted by $f(y_i|y_{\sim i})$, for $i = 1, \ldots, n$, and regression models both for the mean and the dispersion parameters are specified, including covariates and spatial lags of the response variable

Generalized spatial conditional normal Poisson

It assumes that $(Y_i|Y_{\sim i}, \nu_i) \sim \text{Poi}(\mu_i)$, then regression models both for the mean and the dispersion parameters are specified, so that:

$$\begin{aligned} \log(\mu_i) =& \mathbf{X}_i'\boldsymbol{\beta} + \rho_1 \mathbf{W}_i \mathbf{y} + \nu_i \quad \text{with} \quad \nu_i \sim N(0, \tau_i) \quad \text{and} \\ \log(\tau_i) =& \mathbf{Z}_i'\gamma + \rho_2 \mathbf{W}_i \mathbf{y}, \end{aligned}$$

where the parameters ρ_1 and ρ_2 explain the strength of the spatial association in the mean and the dispersion respectively

Bayesian framework

- Data: n independent observations, y_i, for $i = 1, \ldots, n$ from the variable $\mathbf{Y}$
- Parameters of interest: $\theta \Longrightarrow$ Express our belief via a prior distribution $p(\theta)$
- Aim: Make inference about θ using the information available in the data with the likelihood $L(\mathbf{y}|\theta)$
- How? Update our knowledge about θ using the Bayes theorem $\Longrightarrow$ Obtain a posterior distribution for θ: $p(\theta|\mathbf{y}) \propto L(\mathbf{y}|\theta)p(\theta)$
- Very often the posterior is intractable $\Longrightarrow$ Use computational methods such as Markov chain Monte Carlo (MCMC) and Integrated Nested Laplace Approximation (INLA) (Software: JAGS, WinBUGS, R-INLA, Stan, NIMBLE and a variety of R packages)
- No prior information available $\Longrightarrow$ Vague or non informative priors

Study of infant mortality rates in Colombia (2005)

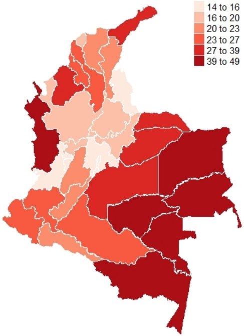
Variables available for each of the $n = 32$ departments (regions) are:

- **ND**: Number of children under one year of age who died in 2005
- **NB**: Total number of births $\Longrightarrow$ $\mathbf{Rates}_i = \frac{\mathrm{ND}_i}{\mathrm{NB}_i} \times 1000, \;\; i = 1, \ldots, n$
- **IBN**: Index representing the percentage of the population with basic needs not satisfactorily attended.
- **Rec**: Resources (in thousands of dollars) provided by the government per household for academic achievement or education and integral attention for children and young people
- **Viol**: Percentage of women over the age of 18 who had suffered physical violence from their current partners
- **HE**: Percentage of people between 18 and 24 years who had access to a higher educational level
- **Vac**: Percentage of children under one year of age who received the third dose of the polio vaccine in the year 2004

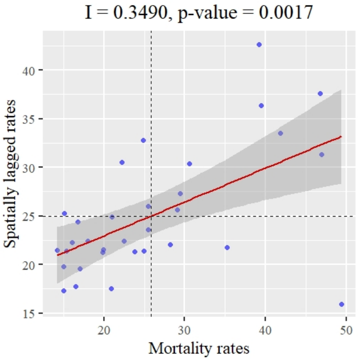
Variable under study

Aim: Study mortality rates, capturing and being able to explain overdispersion and spatial association

Figure: Spatial distribution of the variable Rates



(a) Observed mortality rates



(b) Moran's scatterplot of the variable Rates

Fitted models

- **Poisson**: $\text{ND}_i \sim \text{Poi}(\mu_i)$, with mean following the regression model:

$$\log(\mu_i) = \log(\text{NB}_i) + \beta_0 + \beta_1 \text{Viol}_i + \beta_2 \text{IBN}_i + \beta_3 \text{Rec}_i + \beta_4 \text{HE}_i + \beta_5 \text{Vac}_i$$

- **Spatial conditional Poisson**: $(\text{ND}_i | \text{ND}_{\sim i}) \sim \text{Poi}(\mu_i)$, with mean following the regression model:

$$\begin{aligned}\log(\mu_i) = & \log(\text{NB}_i) + \beta_0 + \beta_1 \text{Viol}_i + \beta_2 \text{IBN}_i + \beta_3 \text{Rec}_i + \beta_4 \text{HE}_i + \beta_5 \text{Vac}_i \\ & + \rho \mathbf{W}_i \mathbf{Rates}\end{aligned}$$

- **Normal Poisson**: $(\text{ND}_i | \nu_i) \sim \text{Poi}(\mu_i)$, with mean following the regression model:

$$\begin{aligned}\log(\mu_i) = & \log(\text{NB}_i) + \beta_0 + \beta_1 \text{Viol}_i + \beta_2 \text{IBN}_i + \beta_3 \text{Rec}_i + \beta_4 \text{HE}_i + \beta_5 \text{Vac}_i \\ & + \nu_i, \quad \text{where} \quad \nu_i \sim N(0, \tau)\end{aligned}$$

Fitted models

- **Spatial conditional normal Poisson**: $(\mathrm{ND}_i | \mathrm{ND}_{\sim i}, \nu_i) \sim \mathrm{Poi}(\mu_i)$, with mean following the regression model:

$$\begin{aligned}\log(\mu_i) = &\log(\mathrm{NB}_i) + \beta_0 + \beta_1 \mathrm{Viol}_i + \beta_2 \mathrm{IBN}_i + \beta_3 \mathrm{Rec}_i + \beta_4 \mathrm{HE}_i + \beta_5 \mathrm{Vac}_i \\ &+ \rho \mathbf{W}_i \mathbf{Rates} + \nu_i, \quad \text{where} \quad \nu_i \sim N(0, \tau)\end{aligned}$$

- **Negative binomial**: $\mathrm{ND}_i \sim \mathrm{NB}(\tau/(\tau + \mu_i), \tau)$ with mean following the regression model:

$$\log(\mu_i) = \log(\mathrm{NB}_i) + \beta_0 + \beta_1 \mathrm{Viol}_i + \beta_2 \mathrm{IBN}_i + \beta_3 \mathrm{Rec}_i + \beta_4 \mathrm{HE}_i + \beta_5 \mathrm{Vac}_i$$

- **Spatial conditional negative binomial**: $(\mathrm{ND}_i | \mathrm{ND}_{\sim i}) \sim \mathrm{NB}(\tau/(\tau + \mu_i), \tau)$ with mean following the regression model:

$$\begin{aligned}\log(\mu_i) = &\log(\mathrm{NB}_i) + \beta_0 + \beta_1 \mathrm{Viol}_i + \beta_2 \mathrm{IBN}_i + \beta_3 \mathrm{Rec}_i + \beta_4 \mathrm{HE}_i + \beta_5 \mathrm{Vac}_i \\ &+ \rho \mathbf{W}_i \mathbf{Rates}\end{aligned}$$

Bayesian estimation

- Specified noninformative prior distributions for all the regression parameters:
  - $\beta_j \sim N(0, 1 \times 10^5), j = 1, \dots, k$ and $\rho \sim N(0, 1 \times 10^5)$
  - $\frac{1}{\tau} \sim G(1 \times 10^{-4}, 1 \times 10^{-4}) \Longrightarrow$ Performed a sensitivity analysis, considering different possible values $G(\alpha, \alpha)$ from $\alpha = 0.1$ to $\alpha = 1 \times 10^{-8}$, to make sure that the prior would not affect posterior inference
- We have used OpenBUGS, JAGS (Markov chain Monte Carlo - MCMC approach) and Integrated Nested Laplace Approximation (INLA)
- Model comparison and selection was carried out by:
  - Deviance Information Criterion (DIC)
  - Watanabe-Akaike Information Criterion (WAIC)

  Lowest values indicate better fit
  - Performing posterior predictive checks

Variable under study: Overdispersion

- **Poisson**: $\text{ND}_i \sim \text{Poi}(\mu_i)$, with mean following the regression model:

$$\log(\mu_i) = \log(\text{NB}_i) + \beta_0 + \beta_1 \text{Viol}_i + \beta_2 \text{IBN}_i + \beta_3 \text{Rec}_i + \beta_4 \text{HE}_i + \beta_5 \text{Vac}_i$$

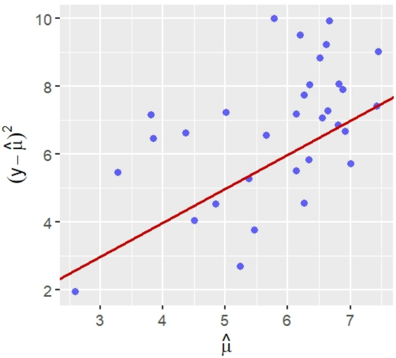
	Mean	SD	95% CI
Intercept	-4.2960	(0.0902)	(-4.4710,-4.1210)
Viol	0.0066	(0.0015)	(0.0036,0.0096)
IBN	0.0155	(6.0763E-04)	(0.0143,0.0167)
Rec	-4.9534E-04	(1.2787E-04)	(-7.4542E-04,-2.4597E-04)
HE	-0.0013	(9.2185E-04)	(-0.0031,4.5622E-04)
Vac	-0.0036	(9.4170E-04)	(-0.0054,-0.0017)

DIC = 491.7, WAIC = 524.0

Variable under study: Overdispersion

No overdispersion $\Longrightarrow$ $\hat{\mu}_i$ should be approximately equal to the estimated variance $(y_i - \hat{\mu}_i)^2$.

Figure: Plot of the estimated variance against the mean



$\Longrightarrow$ Most points are above the red line $\Longrightarrow$ Suggests overdispersion

Variable under study: Overdispersion

- **Normal Poisson**: $(\mathrm{ND}_i|\nu_i) \sim \mathrm{Poi}(\mu_i)$, with mean following the regression model:

$$\log(\mu_i) = \log(\mathrm{NB}_i) + \beta_0 + \beta_1 \mathrm{Viol}_i + \beta_2 \mathrm{IBN}_i + \beta_3 \mathrm{Rec}_i + \beta_4 \mathrm{HE}_i + \beta_5 \mathrm{Vac}_i + \nu_i, \quad \text{where} \quad \nu_i \sim N(0, \tau)$$

	Mean	SD	95% CI
Intercept	-4.4246	(0.3235)	(-5.0640,-3.7730)
Viol	0.0108	(0.0061)	(-0.0012,0.0230)
IBN	0.0154	(0.0025)	(0.0105,0.0202)
Rec	-0.0010	(6.0308E-04)	(-0.0022,1.5423E-04)
HE	-0.0057	(0.0044)	(-0.0143,0.0031)
Vac	-0.0014	(0.0031)	(-0.0075,0.0047)
τ	0.0276	(0.0096)	(0.0133,0.0517)
DIC = 308.0, WAIC = 302.3			

Fitted models: results

Model	DIC	WAIC
Poisson	491.7	524.0
Spatial conditional Poisson	480.4	519.6
Normal Poisson	308.0	302.3
Negative Binomial	362.2	361.8
Spatial conditional Negative Binomial	360.1	360.3
Spatial conditional normal Poisson	307.3	301.3

$\Longrightarrow$ Model with lowest information criteria values: **Spatial conditional normal Poisson**

$\Longrightarrow$ Variable selection taking into account information criteria values and posterior predictive checks

Spatial conditional normal Poisson reduced

- $(\text{ND}_i|\text{ND}_{\sim i}, \nu_i) \sim \text{Poi}(\mu_i)$, with mean following the regression model:

$$\begin{aligned}\log(\mu_i) =& \log(\text{NB}_i) + \beta_0 + \beta_1 \text{IBN}_i + \beta_2 \text{Rec}_i \\ & + \rho \mathbf{W}_i \mathbf{Rates} + \nu_i, \quad \text{where} \quad \nu_i \sim N(0, \tau)\end{aligned}$$

	Mean	SD	95% CI
Intercept	-4.6479	(0.1660)	(-4.9660,-4.3240)
IBN	0.0167	(0.0018)	(0.0132,0.0202)
Rec	-0.0011	(4.9818E-04)	(-0.0021,-1.8238E-04)
ρ	0.0151	(0.0061)	(0.0033,0.0272)
τ	0.0233	(0.0080)	(0.0119,0.0430)
DIC = 307.4, WAIC = 302.3			

- Nonconstant dispersion?

Generalized spatial conditional normal Poisson reduced

Performed variable selection process $\Longrightarrow$ Best model:

$(\mathrm{ND}_i|\mathrm{ND}_{\sim i}, \nu_i) \sim \mathrm{Poi}(\mu_i)$ with mean following the regression model:

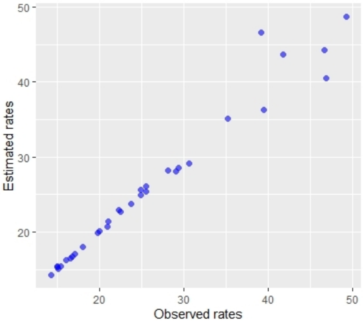
$$\log(\mu_i) = \log(\mathrm{NB}_i) + \beta + \rho \mathbf{W}_i \mathbf{Rates} + \nu_i, \quad \text{where} \quad \nu_i \sim N(0, \tau_i)$$
$$\log(\tau_i) = \gamma_0 + \gamma_1 \mathrm{IBN}_i$$

	Mean	SD	95% CI
μ **Intercept** (β)	-4.9262	(0.2275)	(-5.3620,-4.4780)
ρ	0.0431	(0.0093)	(0.0249,0.0608)
τ **Intercept** (γ_0)	-4.2186	(0.6165)	(-5.4200,-2.9998)
IBN	0.0430	(0.0145)	(0.0161,0.0727)

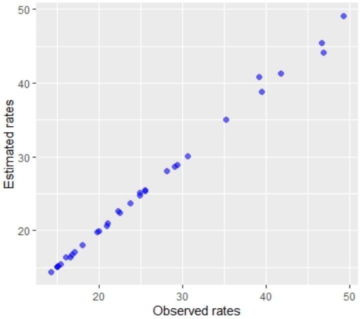
DIC = 308.0, WAIC = 299.1

Posterior predictive checks

Figure: Scatterplots for the observed versus the predicted rates obtained from some of the fitted models to the Colombia infant mortality rates data set



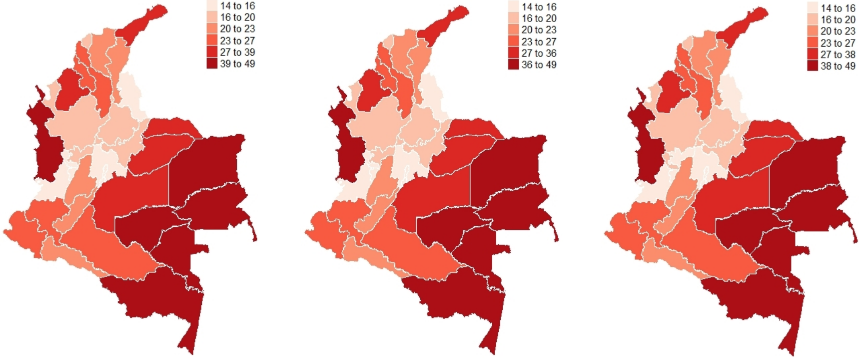
(a) Spatial conditional normal Poisson with variables NBI and Rec



(b) Generalized spatial conditional normal Poisson

Posterior predictive checks

Figure: Maps of the observed and estimated mortality rates obtained from some of the fitted models to the Colombian infant mortality rates data set



(a) Observed rates

(b) Spatial conditional normal Poisson with variable IBN and Rec

(c) Generalized spatial conditional normal Poisson

Comparison to the Besag–York–Mollié (BYM) models

BYM model

Assume that $(Y_i|\nu_i, \eta_i) \sim \text{Poi}(\mu_i)$, with conditional mean $\mathrm{E}(Y_i|\nu_i, \eta_i) = \mu_i$ following the regression model:

$$\log(\mu_i) = \mathbf{X}_i'\boldsymbol{\beta} + \nu_i + \eta_i,$$

where $\nu_i \sim N(0, \tau)$ is an unstructured random effect, and η_i is a spatially structured effect that follows an intrinsic conditional autoregressive (CAR) distribution. $\Longrightarrow$ **Problem: Identifiability**

BYM2 model

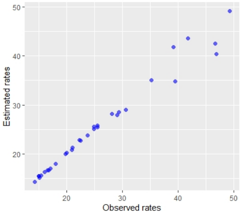
Extension of the BYM model that scales the spatial component and the unstructured component, so that the mean regression structure can be written as:

$$\log(\mu_i) = \mathbf{X}_i'\boldsymbol{\beta} + \frac{1}{\sqrt{\tau_s}}\left(\sqrt{1-\phi_s}\,\nu_i + \sqrt{\phi_s}\,\eta_i\right),$$

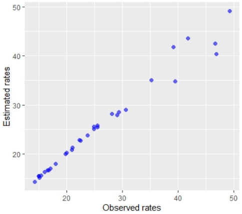
Comparison to the Besag–York–Mollié (BYM) models

- We have fitted the BYM and BYM2 models to the same dataset
- Do not offer improvements in terms of DIC and WAIC when compared to the spatial conditional models
- Do not offer improvements in terms of posterior predictive accuracy

Figure: Scatterplots for the observed versus the predicted rates obtained from some of the fitted BYM and BYM2 models to the Colombia infant mortality rates data set



(a) BYM model with variables IBN and Rec



(b) BYM2 model with variables IBN and Rec

Study of COVID-19 incidence rates in Flanders, Belgium
(Joint work with Prof. Christel Faes from Hasselt University)

Variables available for each of the $n = 300$ municipalities are:

- **Ncases**: Number of COVID-19 in each municipality from September 2020 to January 2021
- **P**: Population in each municipality

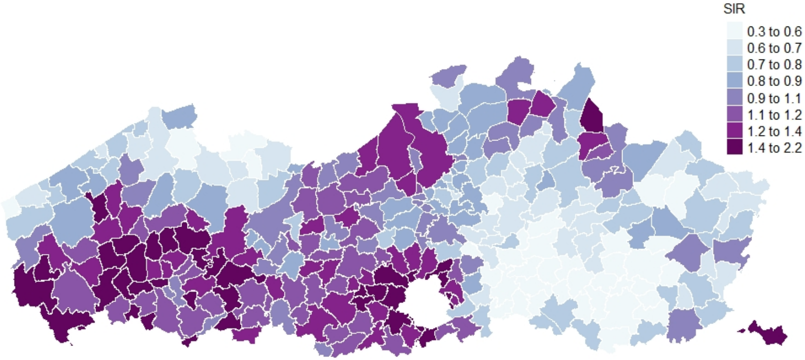
$\Longrightarrow$ $\text{Incidence}_i = \dfrac{\text{Ncases}_i}{\text{P}_i} \times 100000$

$\Longrightarrow$ Standardized Incidence Ratio (SIR): $\text{SIR}_i = \dfrac{\text{Ncases}_i}{\text{E}_i}$

- E_i: expected counts if the population in the specific region behaved as the standard or overall population does $\Longrightarrow$ Using indirect standardization: $\text{E}_i = r^{(s)}\text{P}_i$ (Could also use age and/or gender)
- $r^{(s)}$: rate for the entire population, so that $r^{(s)} = \dfrac{\sum_{i=1}^{n} \text{Ncases}_i}{\sum_{i=1}^{n} \text{P}_i}$

Study of COVID-19 incidence in Flanders, Belgium

Figure: Map of the Standardized Incidence Ratio (SIR)



$\Longrightarrow$ $\text{SIR}_i > 1$ indicates larger incidence than expected in region i and $\text{SIR}_i < 1$ indicates lower

Fitted models

The SIRs do not take into account spatial association and are sensible to extreme values given by small areas $\Longrightarrow$ Estimate relative risks of the disease for each region via a spatial regression model

Spatial conditional normal Poisson for relative risks θ_i

$(\text{Ncases}_i | \text{Ncases}_{\sim i}, \text{E}_i, \nu_i) \sim \text{Poi}(\text{E}_i \theta_i)$, with mean following the regression model:

$$\log(\theta_i) = \beta + \rho \mathbf{W}_i \textbf{Incidence} + \nu_i, \quad \text{where} \quad \nu_i \sim N(0, \tau)$$

Different specifications for $\mathbf{W} = [w_{ij}]$:

- Contiguity of order 1
- Contiguity of order 3
- Inverse distance
- Negative exponential
- Distance band: Minimum distance that ensures all regions have at least one neighbor

Fitted models

Another connectivity structure $\Longrightarrow$ Mobility matrix: $\mathbf{M} = [m_{ij}]$

- m_{ij}: mean proportion of time people from municipality i have spent in municipality j from September to December 2020
- Standardized by rows $\Longrightarrow$ $\mathbf{M}_i\textbf{Incidence}$, can be considered as an average of the incidence in the municipalities where people moved into, weighted by the mean proportion of time they spent there.

Spatial conditional normal Poisson for relative risks θ_i

$(\text{Ncases}_i|\text{Ncases}_{\sim i}, \text{E}_i, \nu_i) \sim \text{Poi}(\text{E}_i\theta_i)$, with mean following the regression model:

$$\log(\theta_i) = \beta + \rho\mathbf{M}_i\textbf{Incidence} + \nu_i, \quad \text{where} \quad \nu_i \sim N(0, \tau)$$

Fitted models: results

Model	DIC	WAIC
Contiguity of order 1	3015.6	2947.7
Contiguity of order 3	3019.1	2942.7
Inverse distance	3017.9	2941.2
Negative exponential	3022.1	2941.4
Distance band	3013.5	2938.4
Mobility	3029.9	2972.7

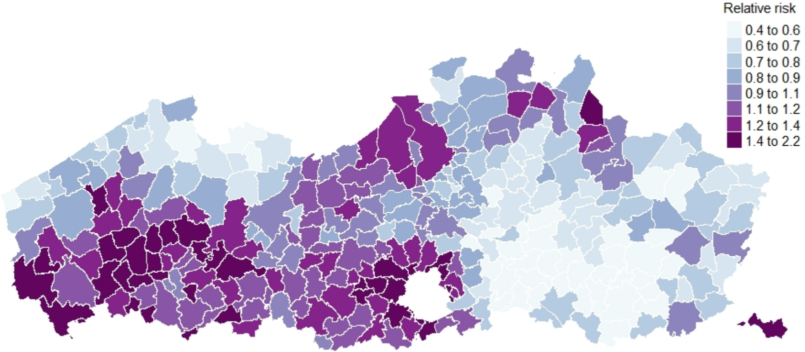
$\Longrightarrow$ Model with the lowest information criteria values: $\mathbf{W}$ following the distance band specification

Fitted models: results

		Contig. of order 1	Distance band	Mobility
β	Mean	-1.0609	-1.1969	-0.9528
	SD	(0.0414)	(0.0505)	(0.0445)
	95% CI	(-1.1423,-0.9797)	(-1.2963,-1.0979)	(-1.0401,-0.8654)
ρ	Mean	27.9277	31.4797	25.1133
	SD	(1.1205)	(1.3702)	(1.2129)
	95% CI	(25.7278,30.1286)	(28.7900,34.1715)	(22.7264,27.4908)
τ	Mean	0.0362	0.0409	0.0458
	SD	(0.0032)	(0.0036)	(0.0041)
	95% CI	(0.0304 0.0430)	(0.0344 0.0484)	(0.0383,0.0546)
	DIC	3015.6	3013.5	3029.9
	WAIC	2947.7	2938.4	2972.7

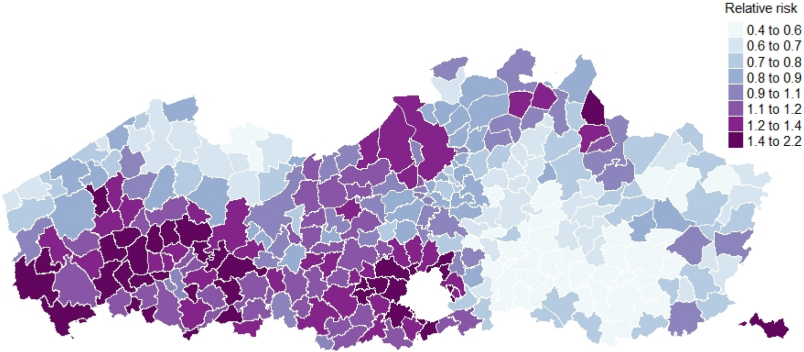
Fitted models: results

Figure: Estimated relative risks obtained from the model with spatial weights following **contiguity of order 1**



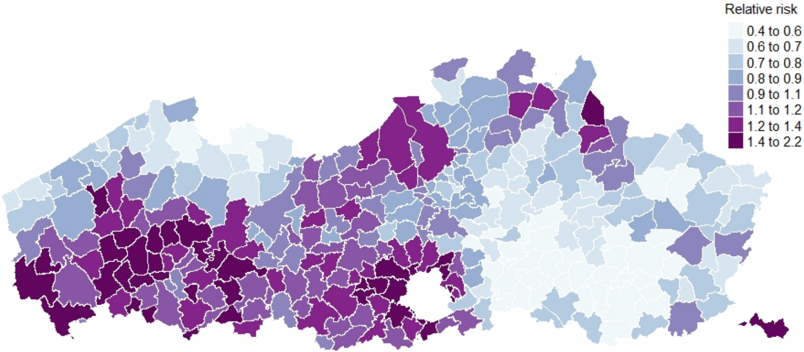
Fitted models: results

Figure: Estimated relative risks obtained from the model with spatial weights following **distance band**



Fitted models: results

Figure: Estimated relative risks obtained from the model with spatial weights following the **mobility** connectivity structure



Fitted models: results

Exceedance probabilities

$P(\theta_i > c) = 1 - P(\theta_i \leq c)$: Probability of the estimated relative risk of a region being greater than a given value c

$\Longrightarrow$ Indicate how likely the risk is to exceed the value c

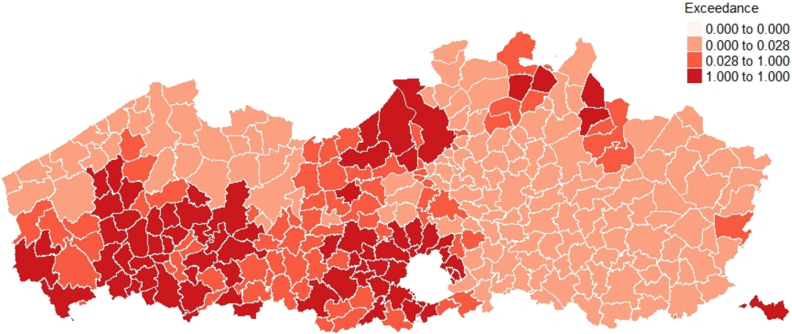
$\Longrightarrow$ Identify areas with an unusual elevated risk

$\Longrightarrow$ Which regions should be examined more closely

$\Longrightarrow$ Where should public health policies be applied

Fitted models: results

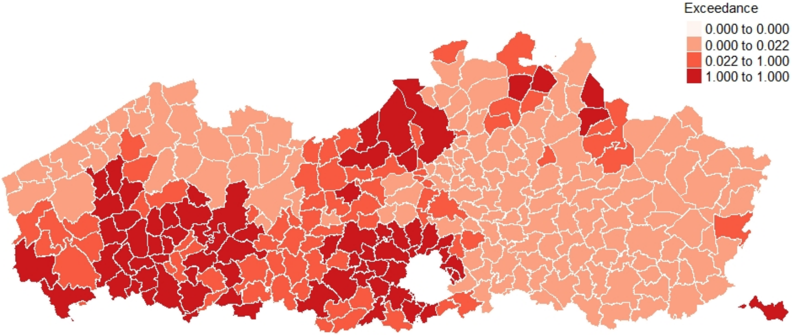
Figure: Exceedance probabilities for $c = 1$ ($\mathrm{P}(\theta_i > 1)$), obtained from the model with spatial weights following **contiguity of order 1**



$\Longrightarrow$ In darker regions, the risk is very likely to exceed 1

Fitted models: results

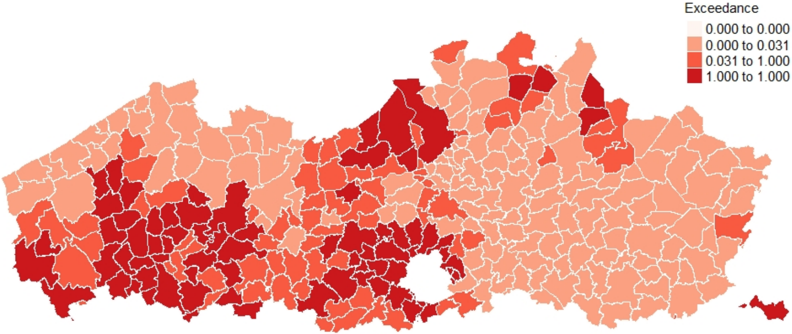
Figure: Exceedance probabilities for $c = 1$ ($\mathrm{P}(\theta_i > 1)$), obtained from the model with spatial weights following **distance band**



$\Longrightarrow$ In darker regions, the risk is very likely to exceed 1

Fitted models: results

Figure: Exceedance probabilities for $c = 1$ ($\mathrm{P}(\theta_i > 1)$), obtained from the model with spatial weights following the **mobility** connectivity structure



$\Longrightarrow$ In darker regions, the risk is very likely to exceed 1

Currently working on:

- Spatial conditional overdispersion models for **Binomial** distributed responses $\Longrightarrow$ Generalized extensions
- **Temporal** extensions of the spatial conditional overdispersion models for spatio-temporal count data $\Longrightarrow$ Following the same idea of Cepeda-Cuervo et al. (2018), but for temporal autocorrelation
- With Prof. **María Durbán** (University Carlos III of Madrid): **Semiparametric** extensions that allow for nonlinear relations among the response and covariates $\Longrightarrow$ Using P-splines as mixed models

Currently working on:

- With Prof. **Christel Faes**: In the context of spatial modelling, constructing **weights matrices** that are not based on contiguity or distance, but instead, reflect **socio-economic characteristics** of similar regions
  $\Longrightarrow$ Based on differences among the values of covariates
  $\Longrightarrow$ Interactions with the standard spatial weights matrix
  $\Longrightarrow$ Application to COVID-19 incidence data in Flanders municipalities

- With Prof. **Virgilio Gómez-Rubio** (University of Castilla La Mancha): Generalized overdispersion models cannot be fitted in INLA
  $\Longrightarrow$ Fitting Double Hierarchical Generalized Linear models in INLA with the Importance Sampling algorithm

Thank you very much for your attention!

mabel.morales@ehu.eus