Politecnico di Milano
Bayesian Statistics

Bayesian Estimation of a spatial-lag autocorrelation model: the Poisson and ZIP cases

Group members

Laura Gamba
Paolo Gerosa
Asia Salpa
Matteo Tomasetto
Sara Tonazzi

Tutor

Prof.ssa Ilenia Epifani

Contents

1 Problem Overview

We are interested in the diffusion of conflicts in 2015 in a region near the lake Chad in Africa; we start from the Master thesis of Galli [2]. As we can see from the map below, in Figure 1, the domain of interest is subdivided in 505 cells where we observe five covariates and the target variable for each region.

In particular, for each cell there are:

- *Outcome*: the target variable is the total number of conflicts observed in 2015
- *Covariates*: urbanization, number of ethnic groups, harvested area per capita, water scarcity months and conflicts in 2014 are exogenous variables.

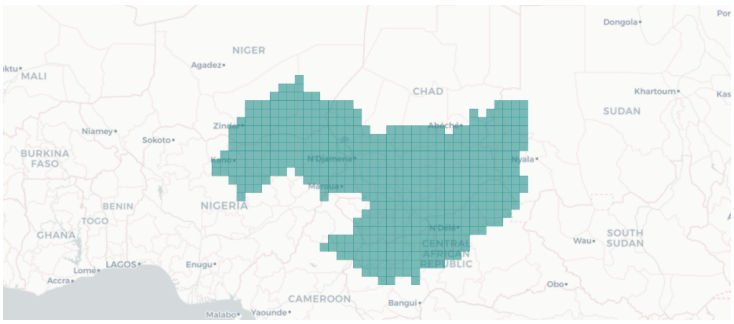


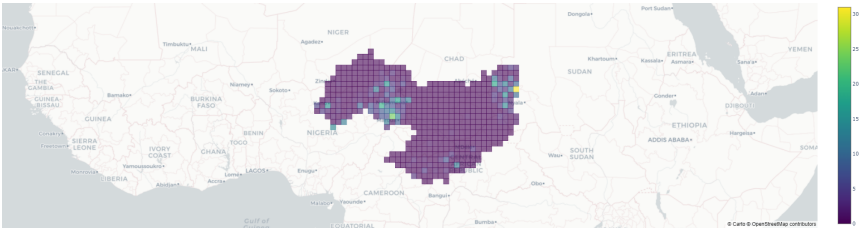
Figure 1: Map of the cells of our problem

The *goal* of our project is to model the impact of the exogenous variables on the number of conflicts in 2015 observed for each cell; to achieve our aim we will consider some different models concerning the following aspects:

- the spatial correlation between the features given by the covariates;
- the spatial correlation of the response variable;
- the massive presence of zeros in the target variable.

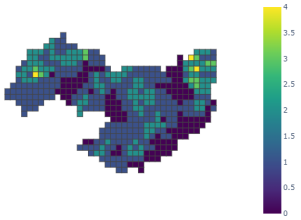
In other words, we will include the effects of the neighboring cells in our model since they might be very meaningful to explain the target.

In order to investigate deeply these peculiarities and to understand if they are relevant with respect to our problem, we perform some preliminary data exploration.



(a) Map of number of conflicts in 2015

(b) Map of water scarcity months



(c) Map of number of ethnic groups

Figure 2

1.1 Data Exploration

First of all, we would like to visualize and assess the spatial variability of the variables and the spatial dependency structure to justify their inclusion in our model.

As we can notice from the maps above in Figure 2, both the outcome and the exogenous variables differ a lot throughout the space; furthermore we can see the presence of *near* regions with similar values: this fact is an indication of spatial correlation, that is of the dependence of the value in a region by the ones of the neighbors.

To quantify this fact, we report in Table 1 the *Moran's index* and the *Moran's p-value* both for covariates and target variable: this is a measure of the global spatial autocorrelation for each variable. In particular, the more the index is near one the more autocorrelation is present for that variable.

Since the p-values are very small, we have a strong frequentist evidence to assess that each variable is strictly spatially autocorrelated; therefore we will consider the spatial correlation effects of covariates and outcome in our model.

In addition, we would like to show two additional aspects that are useful to guide and justify the model choice: firstly, we will have to take into account the wide amount of zeros in the target when choosing the likelihood; indeed

Legend	Variable	Moran's Index	p-value
con_2015	Conflicts in 2015	0.348294	0.001
wsm_2015	Water Scarcity months	0.945106	0.001
hap_2015	Harvested area per capita	0.535472	0.001
etn_2015	Number of ethnic groups	0.550017	0.001
urb_2015	Urbanization	0.111096	0.002

Table 1: Table with *Moran's index* and *Moran's p-value* for each variable

the 75,25% of regions present zero number of conflicts in 2015 (Figure 3 provides the empirical distribution of the conflicts number)

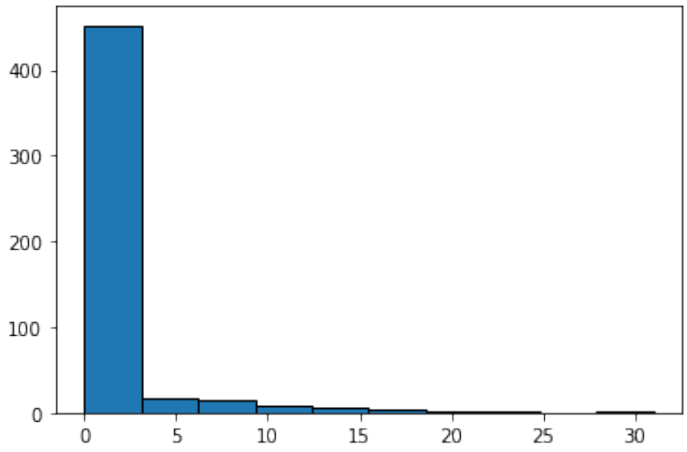


Figure 3: Histogram of the number of conflicts in 2015

Secondly, we want to investigate the correlation and dependency structure between the covariates and the response variable in order to see if it worths to include the exogenous variables in our model:

From the heatmap in Figure 4, we can notice high values of correlation between the target and some covariates. Moreover, thanks to the following plot, we can see the variability of features' mean and features' standard deviation with respect to different bins of the outcome: these results suggest that the covariates have to be considered to explain our response and achieve our goal.

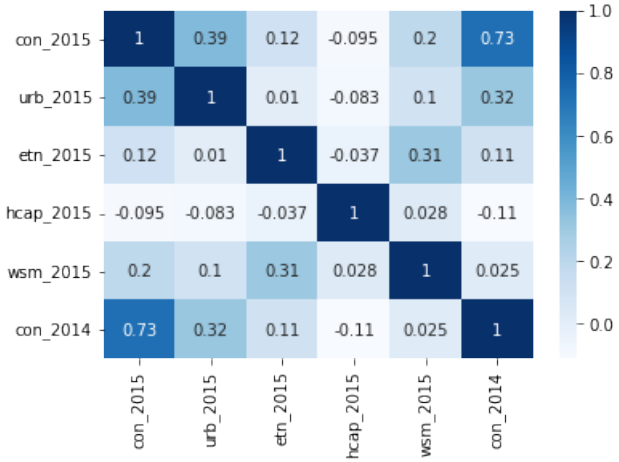


Figure 4: Heatmap of the covariates and the response variable

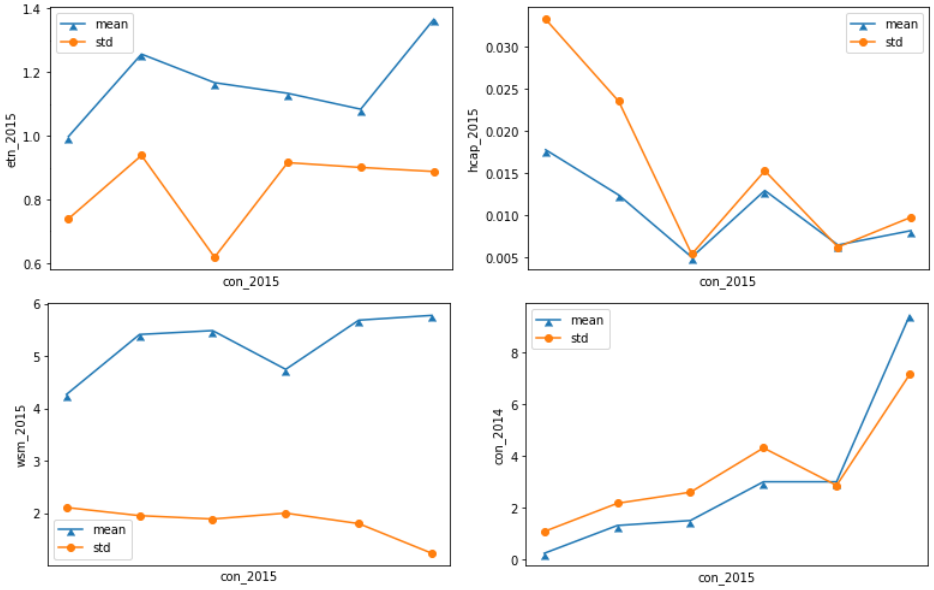


Figure 5: Mean and standard deviation for each covariate grouped by the outcome values

2 Model

Let $N = 505$ and $k = 5$ be the number of cells and the number of covariates; denote by Y_i the number of conflicts in 2015 in the ith cell of our domain. To build a proper *econometric model* we have to take into account two main factors observed in the data exploration:

- the spatial correlation between the observed quantities

- the massive presence of zero in the target variable

We decide to make use of a *log-Poisson regression model* since our target is a count variable: to take into account the spatial correlation, we model the logarithm of the Poisson intensity $\boldsymbol{\mu} = (\mu_1, ..., \mu_N)$ through the data and the responses in neighboring cells. In particular, each $\ln \mu_i$ is built through four ingredients:

1. k exogenous variables (covariates) $X_1, \dots X_k$
2. the means of the k covariates in the neighbouring cells[1]
3. the mean number of the conflicts Y in neighboring cells[2]
4. a random error in order to give more flexibility to the model

Following [3] and [4], we build a *multiplicative Poisson spatial autoregressive lag model for counts*, that we indicate with the acronym P-SAR. This model is part of the models called *Spatial Durbin Models* in the literature.

2.1 Poisson Spatial Durbin Model

Our Poisson Spatial Durbin Model is specified by the following equations:

$$Y_i|\mu_i \overset{\text{ind}}{\sim} \text{Poisson}(\mu_i) \qquad \forall i = 1, ..., N \qquad (1)$$

$$\ln \mu_i = \sum_{j=1}^{5} x_j^{(i)} \beta_j + \sum_{j=1}^{5} \overline{x}_j^{(i)} \gamma_j + \lambda \overline{\ln \mu}^{(i)} + \varepsilon_i \qquad \forall i = 1, ..., N \qquad (2)$$

$$\ln \mu_i = \mathbb{X}_{i\cdot} \boldsymbol{\beta} + [W\mathbb{X}]_{i\cdot} \boldsymbol{\gamma} + \lambda W_{i\cdot} \mathbf{ln}\, \boldsymbol{\mu} + \varepsilon_i \qquad \forall i = 1, ..., N \qquad (3)$$

where:

- $x_j^{(i)}$ is the jth covariates in cell i;
- $\overline{x}_j^{(i)}$ is the mean of the jth covariate in the cells adjacent to the ith cell;
- $\overline{\ln \mu}^{(i)}$ is the mean of the logarithm of the Poisson intensity in the cells adjacent to the ith cell;
- $\mathbb{X}$ is the design matrix;
- $\mathbb{X}_{i\cdot}$ is the ith row of $\mathbb{X}$;
- W is a $N \times N$ row-standardized spatial weights matrix built in order to get the means of covariates and responses in neighboring cells; in particular: $w_{ij} \neq 0$ only if cells i and j share a common border and

[1] Spatial-lag X (SLX) - model
[2] Spatial Autoregressive (SAR) - model

$w_{ij} = 0$ elsewhere. Moreover W is standardized so that all rows sum to one and, by convention, the diagonal elements of W are set to zero;

- $[W\mathbb{X}]_{i\cdot}$ is the ith row of $W\mathbb{X}$ and provides the vector of the means of each covariate $X_1, \ldots, X_k$ in the cells adjacent to the ith cell;
- $W_{i\cdot}\mathbf{ln}\,\boldsymbol{\mu}$ is the mean of $\ln \mu_1, ..., \ln \mu_N$ restricted to the neighbors of the ith cell;
- $\boldsymbol{\beta}$ is the vector of fixed effects coefficients;
- $\boldsymbol{\gamma}$ is the vector of spillovers;
- λ is the spatial autocorrelation parameter.

Equations (2) and (3) are two alternative ways to represent each $\ln \mu_i$. We can write model (1), (2) also in the following matrix form:

$$\mathbf{ln}\,\boldsymbol{\mu} = (I - \lambda W)^{-1}(\mathbb{X}\boldsymbol{\beta} + W\mathbb{X}\boldsymbol{\gamma} + \boldsymbol{\varepsilon}) \tag{4}$$

where $\lambda \in (0, 1)$ guarantees to have the matrix $(I - \lambda W)$ invertible.

To complete our model, we choose some priors for the unknown parameters, commonly used in the literature:

$$\begin{aligned}
&\boldsymbol{\beta} \perp\!\!\!\perp \boldsymbol{\gamma} \perp\!\!\!\perp \lambda \perp\!\!\!\perp \boldsymbol{\varepsilon} \\
&\beta_j|\sigma_\beta^2 \overset{iid}{\sim} \mathcal{N}(0, \sigma_\beta^2) && \forall j = 1, ..., K \\
&\gamma_j|\sigma_\gamma^2 \overset{iid}{\sim} \mathcal{N}(0, \sigma_\gamma^2) && \forall j = 1, ..., K \\
&\text{logit}(\lambda) \sim \mathcal{N}(0, 5^2) \\
&\varepsilon_i|\sigma_\varepsilon^2 \overset{iid}{\sim} \mathcal{N}(0, \sigma_\varepsilon^2) && \forall i = 1, ..., N \\
&\sigma_\beta, \sigma_\gamma, \sigma_\varepsilon \overset{iid}{\sim} \text{Uniform}(0, 100)
\end{aligned}$$

2.2 Spatial Zero Inflated Poisson Durbin Model

To take into account the excess zeros observed in the number of conflicts, we consider the *Zero-Inflated Poisson* (ZIP) model as likelihood for our data in order to assign an additional mass to the zeros with respect to the standard Poisson density. As suggested by [1], the ZIP likelihood is appropriate when the zeros and the non-zeros values are generated by a single process as in our problem; for this reason, we select this density for our counts rather than other possible choices such as the hurdle likelihood. Moreover, for computational reasons, the additional mass θ in zero, introduced by the ZIP, is modelled as a unique parameter for all the cells, i.e. without

covariates effect. In particular the likelihood is:

$$\begin{aligned}
&\mathbb{P}(Y_i = y_i|\theta, \mu_i) = \theta \mathbb{1}_{(y_i=0)} + (1-\theta)\text{Poisson}(y_i; \mu_i) && \forall i = 1, ..., N \\
&\ln \mu_i = \mathbb{X}_{i\cdot}\boldsymbol{\beta} + [W\mathbb{X}]_{i\cdot}\boldsymbol{\gamma} + \lambda W_{i\cdot}\mathbf{ln}\,\boldsymbol{\mu} + \varepsilon_i && \forall i = 1, ..., N
\end{aligned}$$

With the prior system:

$$\begin{aligned}
&\boldsymbol{\beta} \perp\!\!\!\perp \boldsymbol{\gamma} \perp\!\!\!\perp \lambda \perp\!\!\!\perp \boldsymbol{\varepsilon} \\
&\beta_j|\sigma_\beta^2 \overset{iid}{\sim} \mathcal{N}(0, \sigma_\beta^2) && \forall j = 1, ..., K \\
&\gamma_j|\sigma_\gamma^2 \overset{iid}{\sim} \mathcal{N}(0, \sigma_\gamma^2) && \forall j = 1, ..., K \\
&\text{logit}(\lambda) \sim \mathcal{N}(0, 5^2) \\
&\text{logit}(\theta) \sim \mathcal{N}(0, 5^2) \\
&\varepsilon_i|\sigma_\varepsilon^2 \overset{iid}{\sim} \mathcal{N}(0, \sigma_\varepsilon^2) && \forall i = 1, ..., N \\
&\sigma_\beta, \sigma_\gamma, \sigma_\varepsilon \overset{iid}{\sim} \text{Uniform}(0, 100)
\end{aligned}$$

3 Computational Issues

3.1 Inverse Matrix Problem

The presence of an inverse matrix in Equation (4) leads to computational problems and makes the MCMC simulations unfeasible. In order to face this issue, we follow the strategies presented in Joseph M. (2016) where an intermediate prior for $\mathbf{ln}\,\boldsymbol{\mu}$ is added and where it is considered a sparse representation for W.

To achieve the wanted results, we introduce

$$\boldsymbol{\delta} := (I - \lambda W)\mathbf{ln}\,\boldsymbol{\mu}$$

and rewrite (4) as:

$$\boldsymbol{\delta}|\boldsymbol{\beta}, \boldsymbol{\gamma}, \lambda, \sigma_\varepsilon^2 = \mathbb{X}\boldsymbol{\beta} + W\mathbb{X}\boldsymbol{\gamma} + \boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbb{X}\boldsymbol{\beta} + W\mathbb{X}\boldsymbol{\gamma}, \sigma_\varepsilon^2 I) \tag{5}$$

Normal distribution in (5) has a proper multivariate density if λ lies between the inverse of the minimum and the inverse of the maximum eigenvalue of the spatial weight matrix W, thanks to the *Brook's Lemma*. Furthermore one can prove that the minimum eigenvalue of Wvalue is negative, whereas its maximum eigenvalue is equal to 1. Hence we easily get to the following prior for $\mathbf{ln}\,\boldsymbol{\mu}$

$$\mathbf{ln}\,\boldsymbol{\mu}|\boldsymbol{\beta}, \boldsymbol{\gamma}, \lambda, \sigma_\varepsilon^2 \sim \mathcal{N}\left(\left(I - \lambda W\right)^{-1}(\mathbb{X}\boldsymbol{\beta} + W\mathbb{X}\boldsymbol{\gamma}\right), \Sigma\right)$$

where

$$\Sigma = [\tau(I - \lambda W)^t(I - \lambda W)]^{-1} \quad \text{with} \quad \tau = \frac{1}{\sigma_\varepsilon^2} \tag{6}$$

Therefore we end up with the following new reparameterization of our model

$$\begin{aligned}
&\phi_i := \ln \mu_i && \forall i = 1, ..., N \\
&Y_i|\phi_i \sim \text{Poisson}(\mathrm{e}^{\phi_i}) && \forall i = 1, ..., N \\
&\boldsymbol{\phi}|\boldsymbol{\beta}, \boldsymbol{\gamma}, \lambda, \tau \sim \mathcal{N}(\left(I - \lambda W\right)^{-1}(\mathbb{X}\boldsymbol{\beta} + W\mathbb{X}\boldsymbol{\gamma}), \Sigma) \\
&\boldsymbol{\beta} \perp\!\!\!\perp \boldsymbol{\gamma} \perp\!\!\!\perp \lambda \perp\!\!\!\perp \tau \\
&\beta_j|\sigma_\beta^2 \overset{iid}{\sim} \mathcal{N}(0, \sigma_\beta^2) && \forall j = 1, ..., K \\
&\gamma_j|\sigma_\gamma^2 \overset{iid}{\sim} \mathcal{N}(0, \sigma_\gamma^2) && \forall j = 1, ..., K \\
&\text{logit}(\lambda) \sim \mathcal{N}(0, 5^2) \\
&\sigma_\beta, \sigma_\gamma, \sqrt{\frac{1}{\tau}} \sim \text{Uniform}(0, 100)
\end{aligned}$$

The new representation (6) of the covariance matrix Σ of $\boldsymbol{\phi}$ allows us to take advantage of the results in [5] that provides an efficient sparse representation of the covariance matrix of Conditionally AutoRegressive CAR spatial random effects.

Now we can implement the new model in STAN where we manually specify the log probability of $\boldsymbol{\phi}$ via the log probability accumulator in STAN: this will increase the computational efficiency avoiding the inverse matrix problem. Indeed we compute the log probability of $\boldsymbol{\phi}$ up to additive constants that are negligible in STAN:

$$\begin{aligned}
\log(p(\boldsymbol{\phi}|\lambda, \tau)) &= -\frac{N}{2}\log(2\pi) + \frac{1}{2}\log(\det(\Sigma^{-1})) - \\
&- \frac{1}{2}\left[\boldsymbol{\phi} - (I - \lambda W)^{-1}(\mathbb{X}\boldsymbol{\beta} + W\mathbb{X}\boldsymbol{\gamma})\right]^t \Sigma^{-1} \left[\boldsymbol{\phi} - (I - \lambda W)^{-1}(\mathbb{X}\boldsymbol{\beta} + W\mathbb{X}\boldsymbol{\gamma})\right] \\
&\propto \frac{1}{2}\log(\det(\Sigma^{-1})) - \\
&- \frac{1}{2}\left[\boldsymbol{\phi} - (I - \lambda W)^{-1}(\mathbb{X}\boldsymbol{\beta} + W\mathbb{X}\boldsymbol{\gamma})\right]^t \Sigma^{-1} \left[\boldsymbol{\phi} - (I - \lambda W)^{-1}(\mathbb{X}\boldsymbol{\beta} + W\mathbb{X}\boldsymbol{\gamma})\right] \\
&= \frac{1}{2}\log(\det(\tau(I - \lambda W)^t(I - \lambda W))) - \frac{\tau}{2}\boldsymbol{\eta}^t\boldsymbol{\eta}
\end{aligned}$$

where

$$\boldsymbol{\eta} = (I - \lambda W)\boldsymbol{\phi} - \mathbb{X}\boldsymbol{\beta} - W\mathbb{X}\boldsymbol{\gamma}$$

Since we have

$$\begin{aligned}\det(\tau(I-\lambda W)^t(I-\lambda W)) &= \tau^N \det((I-\lambda W)^t(I-\lambda W)) \\ &= \tau^N \det((I-\lambda W)^t)\det(I-\lambda W) \\ &= \tau^N (\det(I-\lambda W))^2 \\ &= \tau^N \left[\prod_{i=1}^{N}(1-\lambda\xi_i)\right]^2 \\ &= \tau^N \prod_{i=1}^{N}(1-\lambda\xi_i)^2\end{aligned}$$

where $\{\xi_i\}_{i=1}^N$ are the eigenvalues of W computed once at the beginning of the simulation, then we end up with

$$\begin{aligned}\log(p(\boldsymbol{\phi}|\lambda,\tau)) \propto \frac{N}{2}\log(\tau) + \sum_{i=1}^{N}\log(1-\lambda\xi_i) - \\ - \frac{\tau}{2}\left[(I-\lambda W)\boldsymbol{\phi} - \mathbb{X}\boldsymbol{\beta} - W\mathbb{X}\boldsymbol{\gamma}\right]^t\left[(I-\lambda W)\boldsymbol{\phi} - \mathbb{X}\boldsymbol{\beta} - W\mathbb{X}\boldsymbol{\gamma}\right] \quad (7)\end{aligned}$$

Notice that, if we make use of the log-probability (7) in STAN, we avoid any inverse matrix, making the simulations feasible and fast.

Alternatively, thanks to an efficient computation of the determinant by [6], we could also consider the following expansion of $\det(I-\lambda W)$:

$$\begin{aligned}\tau^N(\det(I-\lambda W))^2 &= \tau^N(\det(D^{-1}D(I-\lambda W)))^2 \\ &= \tau^N(\det(D^{-1}))^2(\det(D-\lambda DW))^2 \\ &= \tau^N\left[\prod_{i=1}^{N}(1-\lambda\tilde{\xi}_i)\prod_{i=1}^{N}d_i\right]^2 \\ &= \tau^N\prod_{i=1}^{N}(1-\lambda\xi_i)^2 d_i^2\end{aligned}$$

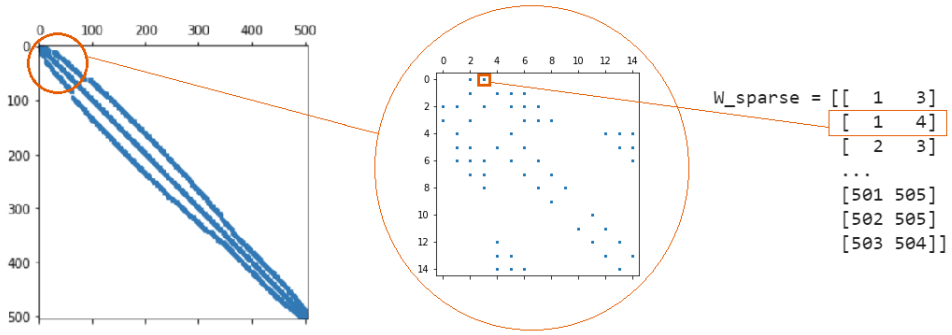
where D is a diagonal matrix with $D_{ii} =$ *number of neighbors of cell* i, while $\{\tilde{\xi}_i\}_{i=1}^N$ and $\{d_i\}_{i=1}^N$ are the eigenvalues of $D^{-\frac{1}{2}}(DW)D^{-\frac{1}{2}}$ and D^{-1} respectively. This new formulation allow us to compute eigenvalues of symmetric matrices only since D^{-1} is diagonal and DW contains 1 if i and j are neighbors and 0 otherwise. In this alternative case we finish up with

$$\log(p(\boldsymbol{\phi}|\lambda,\tau)) \propto \frac{N}{2}\log(\tau) + \sum_{i=1}^{N}\log((1-\lambda\tilde{\xi}_i)d_i) - \frac{\tau}{2}\boldsymbol{\eta}^t\boldsymbol{\eta}$$

3.2 Sparse Model

As a further step, we implement an efficient version of the model using a sparse representation of W in order to increase the computational efficiency; indeed W is a $N \times N$ matrix and it terribly slows down the entire simulation.

Instead of considering the full matrix W, we take into account a sparse formulation where we store the pairs of indices of neighboring cells only:



After this modification, all the computation and matrix products are performed through this sparse version of W; thanks to this trick, we gain in terms of efficiency and run time as we can see from Table 2:

	Initial model	**Modified model**	**Sparse model**
Run time	$+\infty$	1h 11min	14 min

Table 2: Run times

4 Model Checking

We sample from our model with Poisson likelihood using 4 parallel chains and 5000 iterations obtaining the following results:
• Traceplots and posterior distributions for β_1,...,β_5 in Figure 6 along with the values of ESS and the statistic R-hat in Table 3

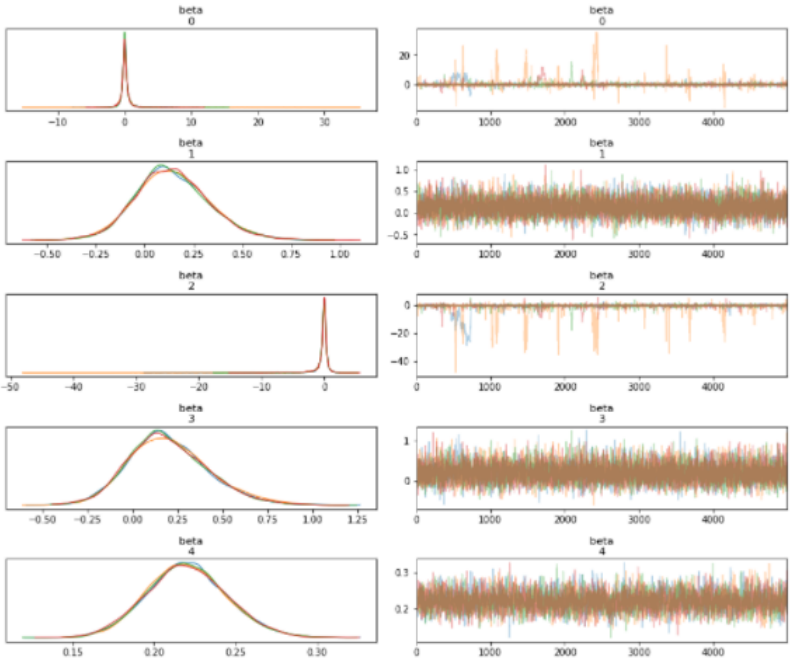
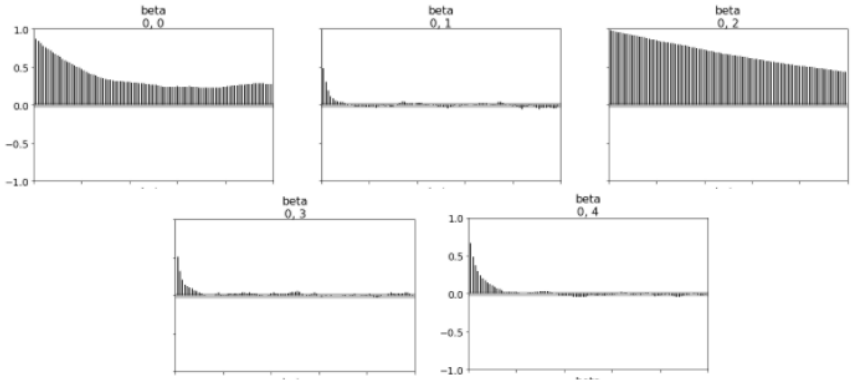


Figure 6: Traceplots and posteriors of β_1,...,β_5

Parameter	**ESS**	**R-hat**
β_0	1560	1.0087
β_1	5432	1.0000
β_2	817	1.0072
β_3	5452	1.0008
β_4	2767	1.0002
γ_0	5632	1.0011
γ_1	2803	1.0005
γ_2	1218	1.0015
γ_3	4409	1.0005
γ_4	2104	1.0016
λ	2472	1.0000
σ^2_β	721	1.0095
σ^2_γ	1173	1.0014
σ^2_ε	1635	1.0013

Table 3: ESS and R-hat for each parameter

- Diverging iterations: 2.68%
- Autocorrelation plot of β_1,...,β_5



After the simulation, we can see that the diverging iterations are quite low; furthermore the values of R-hat and the effective sample sizes are acceptable.
Unfortunately we can easily observe that the posterior for β_0 and β_2 are really skewed; this fact is confirmed by the awful autocorrelation plots for that variables. Similar behaviours are shared by other variables that we omit for simplicity of visualization.
Hence we need to fix these divergence problems increasing the number of iterations or modifying the model: for example we could consider different prior for σ_β, σ_γ and λ. Indeed so far we consider a uniform prior for them but this could provoke issues due to its discontinuous density.

5 Model Development

As first attempt, we try to increase the number of parameters allowing different variances for each β_j and γ_j since they seem to show this feature from the previous traceplots. Hence we end up with the following model:

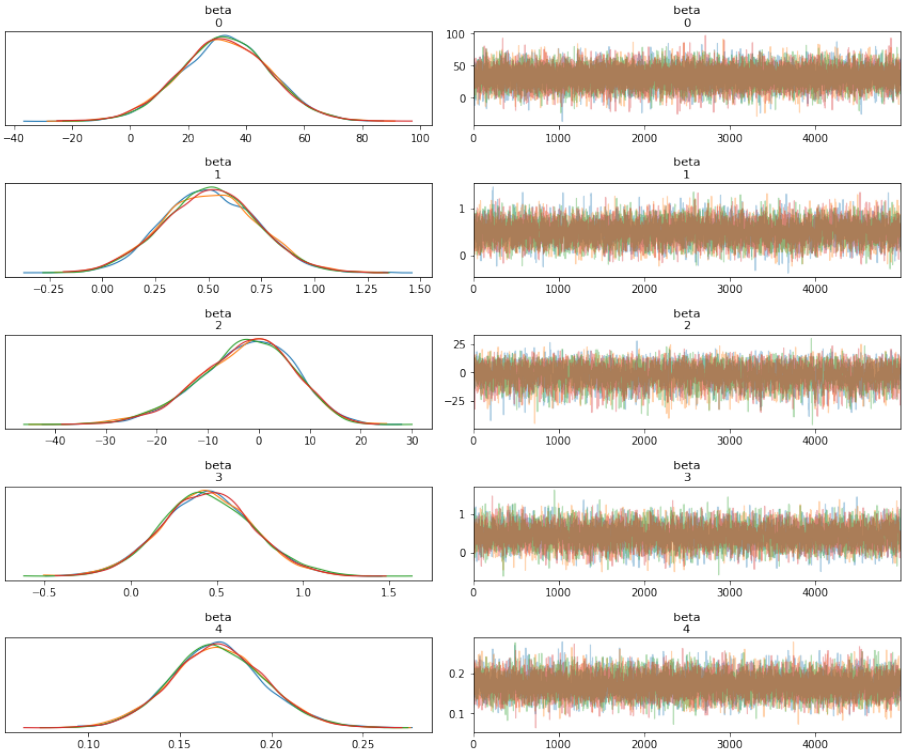
$$\begin{aligned}
&\phi_i := \ln \mu_i && \forall i = 1, ..., N \\
&Y_i | \phi_i \sim \text{Poisson}(\mathrm{e}^{\phi_i}) && \forall i = 1, ..., N \\
&\boldsymbol{\phi} | \boldsymbol{\beta}, \boldsymbol{\gamma}, \lambda, \tau \sim \mathcal{N}((I - \lambda W)^{-1}(\mathbb{X}\boldsymbol{\beta} + W\mathbb{X}\boldsymbol{\gamma}), \Sigma) \\
&\boldsymbol{\beta} \perp\!\!\!\perp \boldsymbol{\gamma} \perp\!\!\!\perp \lambda \perp\!\!\!\perp \tau \\
&\beta_j | \sigma^2_{\beta_j} \overset{iid}{\sim} \mathcal{N}(0, \sigma^2_{\beta_j}) && \forall j = 1, ..., K \\
&\gamma_j | \sigma^2_{\gamma_j} \overset{iid}{\sim} \mathcal{N}(0, \sigma^2_{\gamma_j}) && \forall j = 1, ..., K \\
&\text{logit}(\lambda) \sim \mathcal{N}(0, 5^2) \\
&\sigma_{\beta_j}, \sigma_{\gamma_j}, \sqrt{\frac{1}{\tau}} \sim \text{Uniform}(0, 100) && \forall j = 1, ..., K
\end{aligned}$$

Unfortunately this choice does not give acceptable results: the runtime slows down and make more difficult the simulations with our resources. Moreover we end up with a huge amount of diverging iterations and bad autocorrelation plots.

For this reason, we decide to go in an opposite direction and we simplify the model: in particular we assume deterministic standard deviations $\sigma_\beta = \sigma_\gamma = 50$ and our model reduces in the following way:

$$
\begin{aligned}
&\phi_i := \ln \mu_i && \forall i = 1, ..., N \\
&Y_i | \phi_i \sim \text{Poisson}(\mathrm{e}^{\phi_i}) && \forall i = 1, ..., N \\
&\boldsymbol{\phi} | \boldsymbol{\beta}, \boldsymbol{\gamma}, \lambda, \tau \sim \mathcal{N}\left((I - \lambda W)^{-1} (\mathbb{X}\boldsymbol{\beta} + W\mathbb{X}\boldsymbol{\gamma}) \, , \Sigma \right) \\
&\boldsymbol{\beta} \perp\!\!\!\perp \boldsymbol{\gamma} \perp\!\!\!\perp \lambda \perp\!\!\!\perp \tau \\
&\beta_j \overset{iid}{\sim} \mathcal{N}(0, 50^2) && \forall j = 1, ..., K \\
&\gamma_j \overset{iid}{\sim} \mathcal{N}(0, 50^2) && \forall j = 1, ..., K \\
&\text{logit}(\lambda) \sim \mathcal{N}(0, 5^2) \\
&\sqrt{\frac{1}{\tau}} \sim \text{Uniform}(0, 100)
\end{aligned}
$$

Now the runtime is completely feasible and the results significantly improve as we can see from the traceplots and autocorrelation plots below in Figure 7



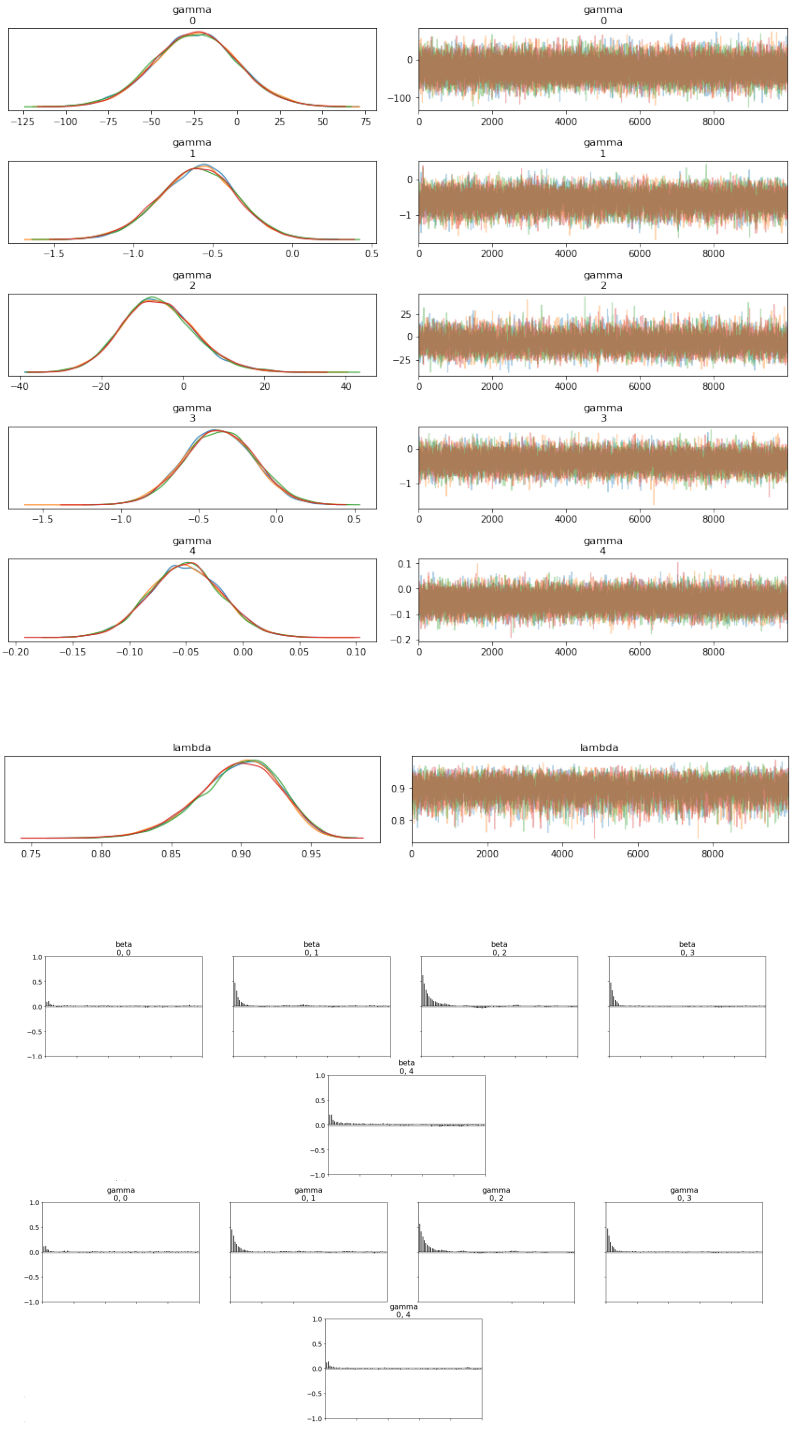


Figure 7: Posterior densities, traceplots and autocorrelation plots of the fixed ($\boldsymbol{\beta}$) and spillover ($\boldsymbol{\gamma}$) effects

We would like to mention another attempt that we have implemented in our work: since the uniform density is discontinuous and may provoke issues in STAN, we pick smoother priors for the variances $\sigma^2_\beta, \sigma^2_\gamma$ such as the inverse-gamma distribution and perform the following

$$
\begin{aligned}
&\phi_i := \ln \mu_i && \forall i = 1, ..., N \\
&Y_i|\phi_i \overset{\text{ind}}{\sim} \text{Poisson}(\text{e}^{\phi_i}) && \forall i = 1, ..., N \\
&\boldsymbol{\phi}|\boldsymbol{\beta}, \boldsymbol{\gamma}, \lambda, \tau \sim \mathcal{N}\left(\left(I - \lambda W\right)^{-1}(\mathbb{X}\boldsymbol{\beta} + W\mathbb{X}\boldsymbol{\gamma}\right), \Sigma) \\
&\boldsymbol{\beta} \perp\!\!\!\perp \boldsymbol{\gamma} \perp\!\!\!\perp \lambda \perp\!\!\!\perp \tau \\
&\beta_j|\sigma^2_\beta \overset{iid}{\sim} \mathcal{N}(0, \sigma^2_\beta) && \forall j = 1, ..., K \\
&\gamma_j|\sigma^2_\gamma \overset{iid}{\sim} \mathcal{N}(0, \sigma^2_\gamma) && \forall j = 1, ..., K \\
&\text{logit}(\lambda) \sim \mathcal{N}(0, 5^2) \\
&\sigma^2_\beta, \sigma^2_\gamma \overset{iid}{\sim} \text{InvGamma}(3, 50) \\
&\sqrt{\frac{1}{\tau}} \sim \text{Uniform}(0, 100)
\end{aligned}
$$

This model provides again good results and all the diagnostics are satisfied; nevertheless we decide to keep the previous simpler model with fixed variances since the results (on the other parameters) are the same and no information criteria tells us that there is a difference among them. According to this choice, we proceed with a simpler model that is faster to fit and that does not give problems of convergence.

6 Model Comparisons

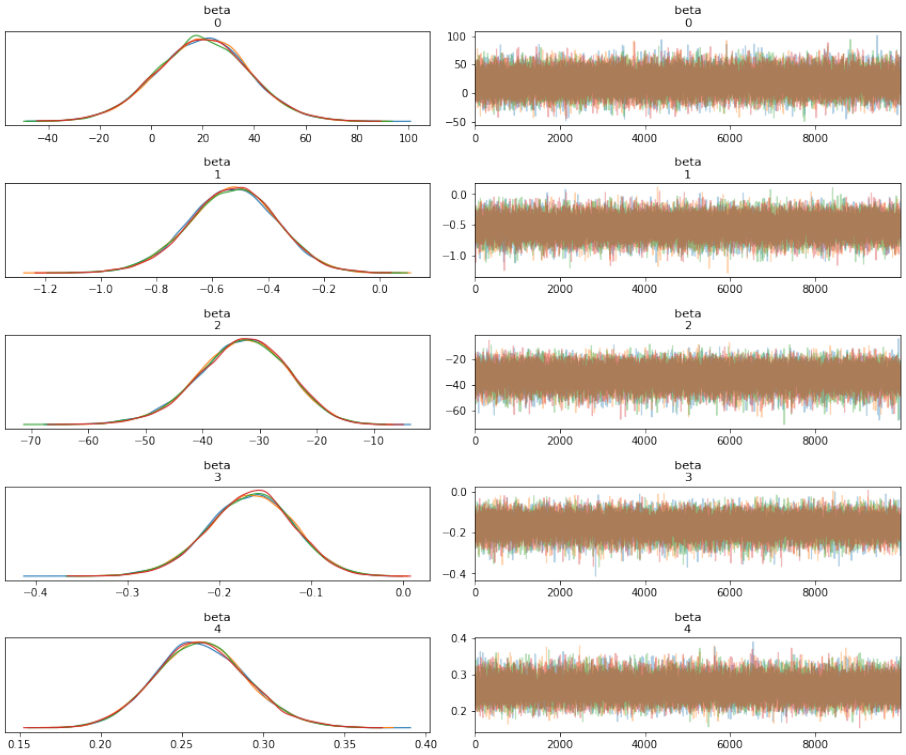
Thanks to the previous steps, we manage to end up with a feasible model that satisfies the diagnostics. In this section, we would like to compare our complete model (with spatial features) to simpler model which do not consider any spatial effects from covariates and from the outcome.

6.1 First Baseline Model

The first baseline model contains only the effect of the covariates, avoiding the contributions of any variables in the neighbors; it is summarized by the following equations:

$$
\begin{aligned}
&Y_i|\mu_i \overset{\text{ind}}{\sim} \text{Poisson}(\mu_i) && \forall i = 1, ..., N \\
&\ln \boldsymbol{\mu} = \mathbb{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \\
&\boldsymbol{\beta} \perp\!\!\!\perp \boldsymbol{\varepsilon} \\
&\beta_j \overset{iid}{\sim} \mathcal{N}(0, 50^2) && \forall j = 1, ..., K \\
&\varepsilon_i|\sigma_\varepsilon^2 \overset{iid}{\sim} \mathcal{N}(0, \sigma_\varepsilon^2) && \forall i = 1, ..., N \\
&\sigma_\varepsilon \sim \text{Uniform}(0, 100)
\end{aligned}
$$

Implementing and fitting this model in STAN, we get the following traceplots and posteriors for the $\boldsymbol{\beta}$ parameters:



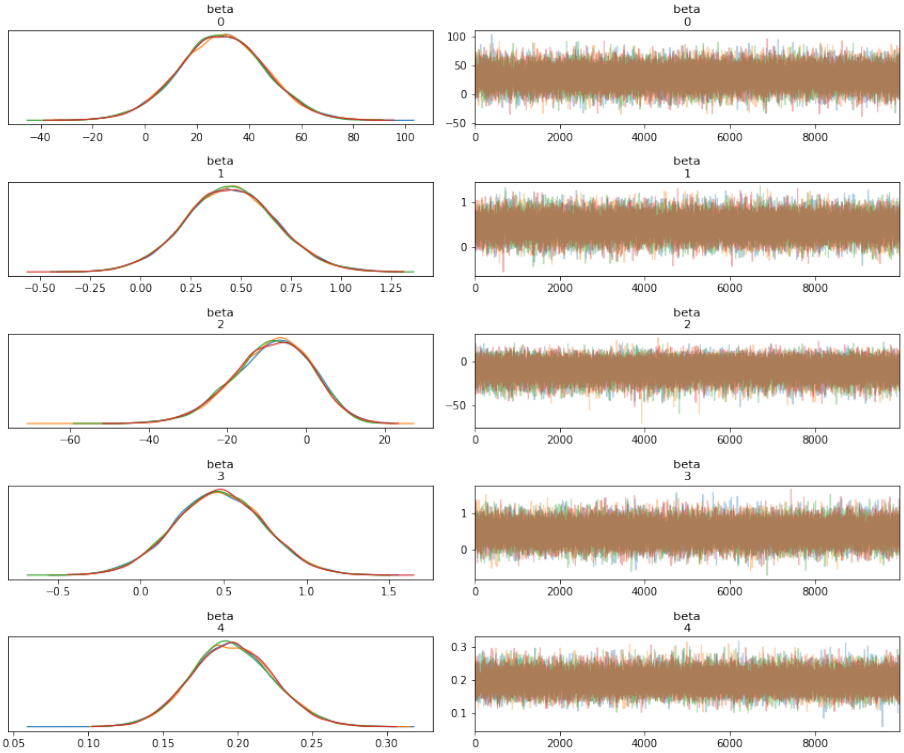
We notice that the traceplots are good without any pathologies; moreover, except for the urbanization (β_0) almost all the covariates are significant to model the target variable since they do not contain the zero value in the 95% credibility interval a posteriori.
Unfortunately some coefficients concentrate on negative values a posteriori: this behaviour is not coherent with the problem. Indeed we expect that variables such as *number of ethnic groups* and *water scarcity months* give positive coefficients in the model increasing the predicted number of conflicts when they assume high values. This fact is highlighted and supported also by the posteriors retrieved in the complete model presented above.
Hence this model is too simple and not enough to model properly our target; for this reason we augment the model considering a more complex baseline case.

6.2 Second Baseline Model

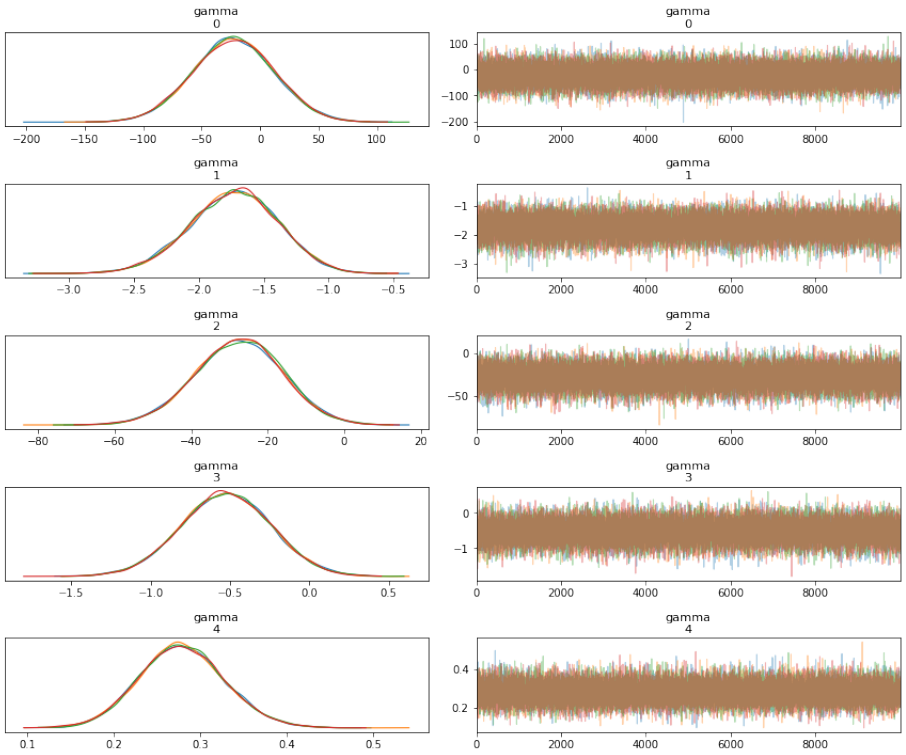
We consider a more complex baseline model adding the SLX term, that is the effects of exogenous variables in the neighboring cells:

$$
\begin{aligned}
&Y_i|\mu_i \overset{\text{ind}}{\sim} \text{Poisson}(\mu_i) && \forall i = 1, ..., N \\
&\mathbf{ln}\, \boldsymbol{\mu} = \mathbb{X}\boldsymbol{\beta} + W\mathbb{X}\boldsymbol{\gamma} + \boldsymbol{\varepsilon} \\
&\boldsymbol{\beta} \perp\!\!\!\perp \boldsymbol{\gamma} \perp\!\!\!\perp \boldsymbol{\varepsilon} \\
&\beta_j \overset{iid}{\sim} \mathcal{N}(0, 50^2) && \forall j = 1, ..., K \\
&\gamma_j \overset{iid}{\sim} \mathcal{N}(0, 50^2) && \forall j = 1, ..., K \\
&\varepsilon_i|\sigma_\varepsilon^2 \overset{iid}{\sim} \mathcal{N}(0, \sigma_\varepsilon^2) && \forall i = 1, ..., N \\
&\sigma_\varepsilon \sim \text{Uniform}(0, 100)
\end{aligned}
$$

Implementing and fitting this model in STAN we get the following traceplots and posteriors for the $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ parameters:



Now the posteriors for the $\boldsymbol{\beta}$ parameters concentrate on positive values as expected. Moreover we get more covariates that are not significant: *urbanization*, *number of ethnic groups*, *harvested area per capita* and *water scarcity months*. This is again quite unexpected since, thanks to our data exploration, all the covariates seemed potentially useful and meaningful to explain the response.
Nevertheless, the $\boldsymbol{\gamma}$ coefficients are significant for our model and so it makes sense to consider the SLX contribution; unfortunately their interpretation is



more difficult now since they link the conflicts in 2015 in a cell with the covariates in the neighbors.
In the complete model presented above, which consider also the spatial correlation of the mean number of conflicts in 2015, we can see that the posterior of λ concentrates on high values; therefore we discover that the autocorrelation parameter λ is extremely meaningful: this suggests that the number of conflicts in a cell strongly depends on the conflicts in the neighboring regions as one can expect. Moreover we can notice that, in the complete model, we have less $\boldsymbol{\gamma}$ parameters that are significant: indeed, considering the neighbouring effects directly through the conflicts in 2015 makes the covariates in the adjacent regions less important to model the response variable.
In conclusion, the full model presents a neighboring effect that is more complete, more interpretable and more compatible with our problem: hence we prefer that model rather than this second baseline case.
This model selection is also supported and guided by the following ranking, based on the LOO error: the lower LOO error in absolute value the better:

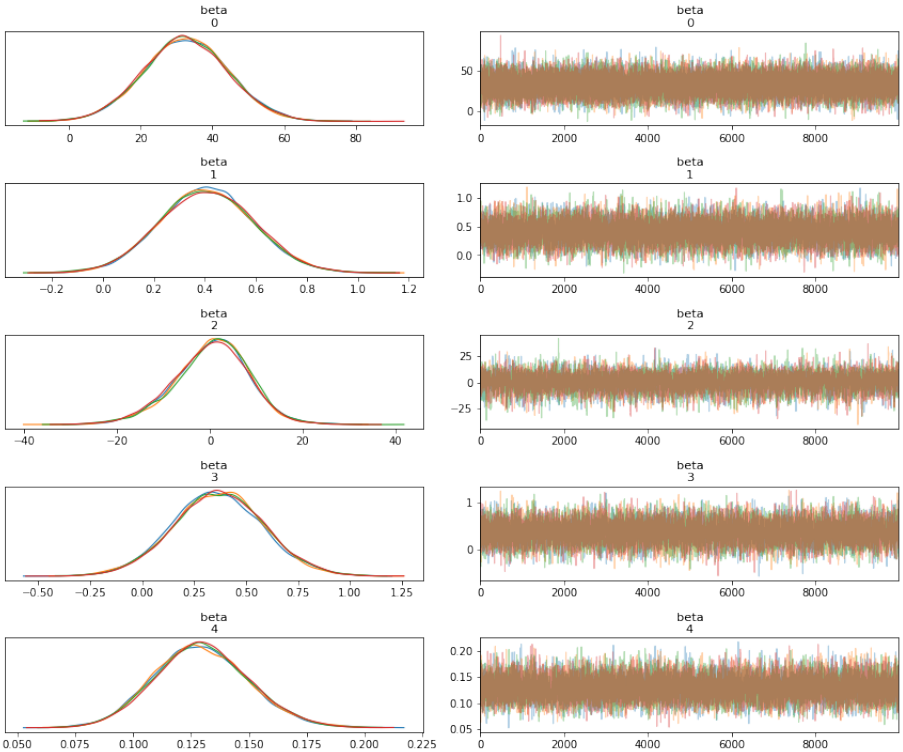
	rank	loo
Complete Model	0	-423.098943
Baseline2	1	-495.577037
Baseline1	2	-509.359024

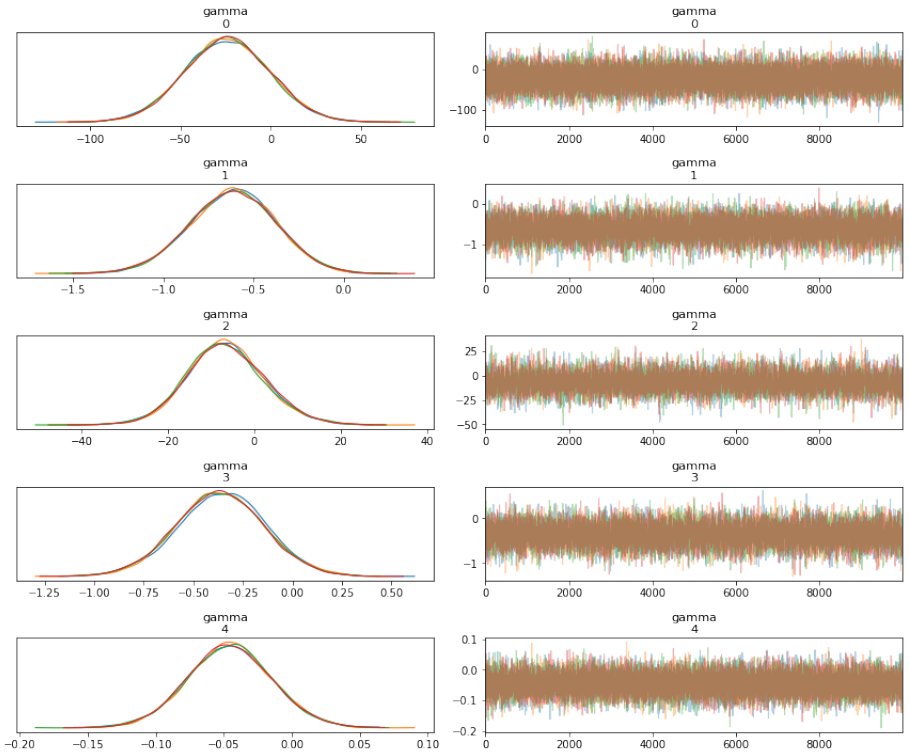
6.3 ZIP Simple Model

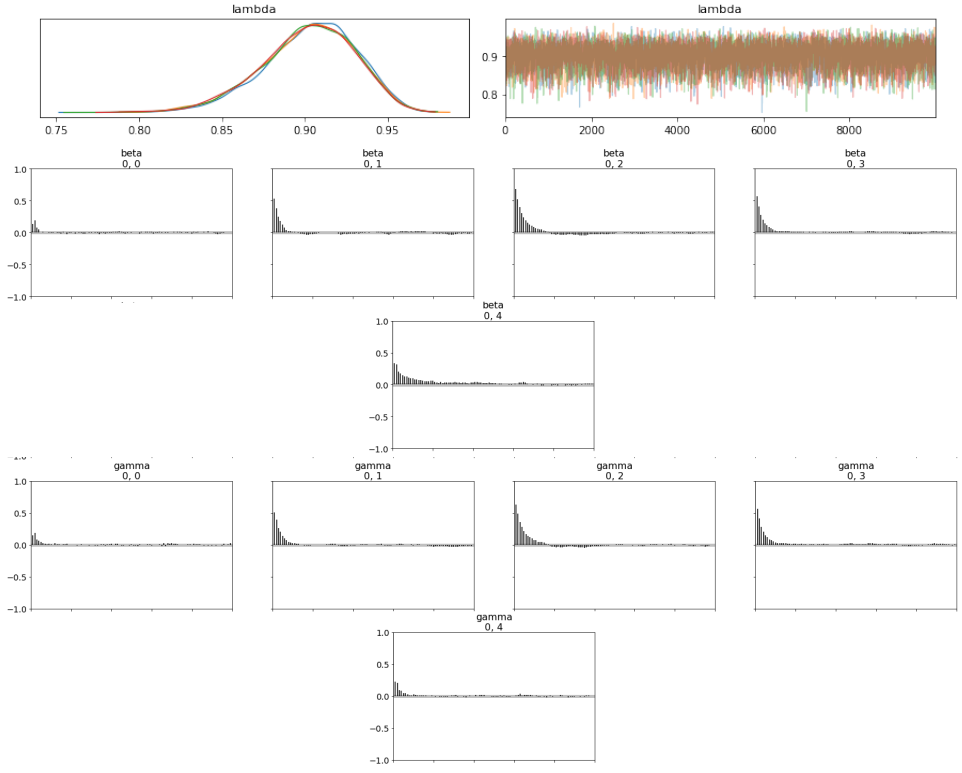
As final comparison, we compare our simple and sparse model presented above with different likelihood: *Poisson* and *Zero Inflated Poisson*. Notice that the ZIP formulation with fixed variances is the following:

$$
\begin{aligned}
&\mathbb{P}(Y_i = y_i|\theta, \mu_i) = \theta \mathbb{1}_{(y_i=0)} + (1-\theta)\text{Poisson}(y_i;\mu_i) && \forall i = 1, ..., N \\
&\ln \mu_i = \mathbb{X}_{i.}\boldsymbol{\beta} + [W\mathbb{X}]_{i.}\boldsymbol{\gamma} + \lambda W_{i.}\mathbf{ln}\,\boldsymbol{\mu} + \varepsilon_i && \forall i = 1, ..., N \\
&\boldsymbol{\beta} \perp\!\!\!\perp \boldsymbol{\gamma} \perp\!\!\!\perp \lambda \perp\!\!\!\perp \boldsymbol{\varepsilon} \\
&\beta_j \overset{iid}{\sim} \mathcal{N}(0, 50^2) && \forall j = 1, ..., K \\
&\gamma_j \overset{iid}{\sim} \mathcal{N}(0, 50^2) && \forall j = 1, ..., K \\
&\text{logit}(\lambda) \sim \mathcal{N}(0, 5^2) \\
&\text{logit}(\theta) \sim \mathcal{N}(0, 5^2) \\
&\varepsilon_i|\sigma_\varepsilon^2 \overset{iid}{\sim} \mathcal{N}(0, \sigma_\varepsilon^2) && \forall i = 1, ..., N \\
&\sigma_\varepsilon \sim \text{Uniform}(0, 100)
\end{aligned}
$$

For the ZIP model fitted in STAN, we have the following traceplots and autocorrelation plots of the fixed ($\boldsymbol{\beta}$) and spillover ($\boldsymbol{\gamma}$) effects.




gamma
0
gamma
0
gamma
1
gamma
1
gamma
2
gamma
2
gamma
3
gamma
3
gamma
4
gamma
4


lambda
lambda
beta
0, 0
beta
0, 1
beta
0, 2
beta
0, 3
beta
0, 4
gamma
0, 0
gamma
0, 1
gamma
0, 2
gamma
0, 3
gamma
0, 4

The complete models with different likelihood seem to be very similar with respect to the posteriors estimates of the parameters (which are quite similar). Nevertheless the ZIP model is preferred since it present a slightly lower *WAIC* index (WAIC = -402) with respect to the Poisson model (WAIC = -379).

7 Posterior Predictive Distribution

As final analysis, we investigate if the posterior predictive distribution of our estimated Poisson ZIP model follows the real data:

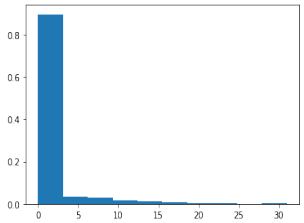


Figure 8: Histogram of the number of conflicts in 2015

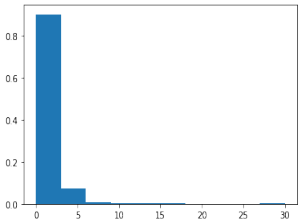


Figure 9: Histogram of predicted outcomes with PSAR model

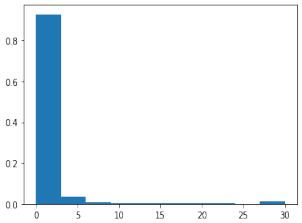


Figure 10: Histogram of predicted outcomes with ZIP model

As we can see from the histograms, our model mimics properly and faithfully the real data giving a posteriori predictive distribution that explains very well our target. Moreover we can see this behaviour also on the maps in Figure 11 and 12:

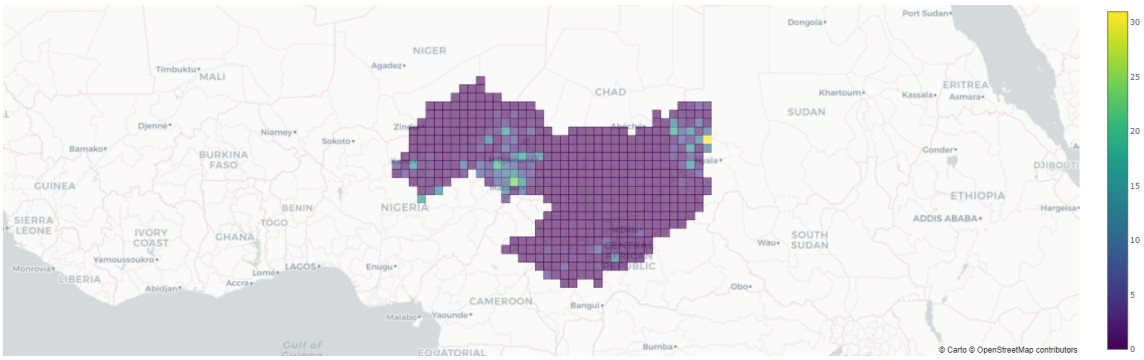


Figure 11: Map of real outcomes

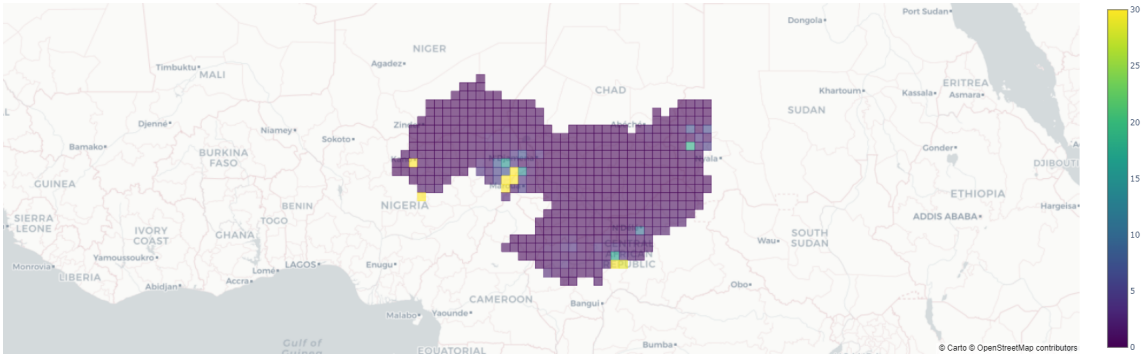


Figure 12: Map of the mean predicted outcomes according to the spatial Poisson ZIP model

Indeed the predicted data on average represent well the real one as observed before; in particular we notice that the spatial structure of the data is replicated very well, thanks to our spatial autocorrelation term in our model.
In addition we exploit the variability of the posterior predictive distribution considering maps where we add and subtract one standard deviation to the mean value of the predicted outcomes for each cell:

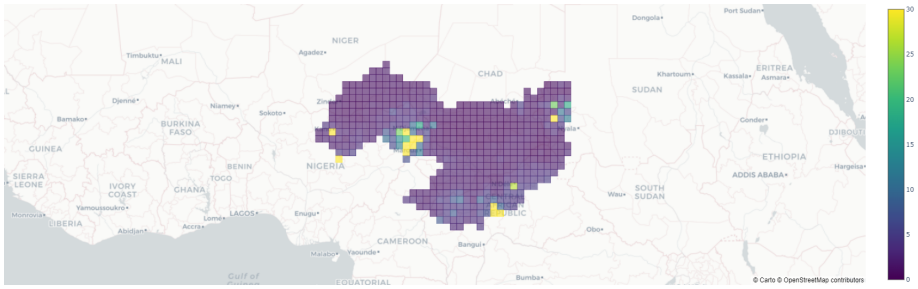


Figure 13: Map of the mean predicted outcomes plus a standard deviation

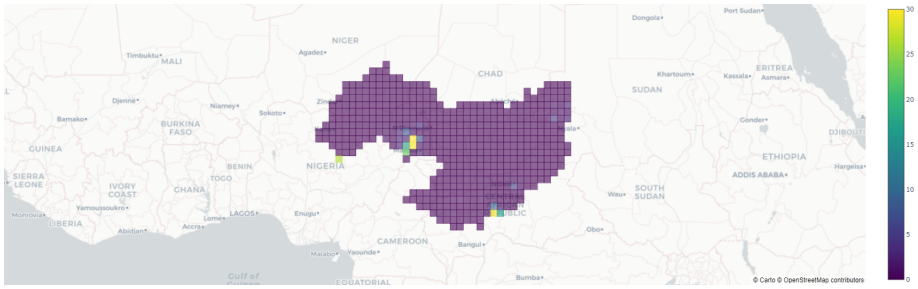


Figure 14: Map of the mean predicted outcomes minus a standard deviation

We can see that the plots are again coherent and remain faithful to the real data with a good precision; moreover the variability of the prediction is linked

with the spatial spread of the outcome: therefore our model properly learns and explains the spatial structure of the conflicts in 2015.

8 Conclusion

The whole analysis strongly suggests that the spatial autocorrelation is crucial to model properly our variable of interest: indeed models with neighboring effects perform better then simpler ones.
Looking at the posterior of the complete model, we discover that *urbanization* and *harvested area per capita* are not significant to model the number of conflicts in the same cell; on the other hand *number of ethnic groups*, *water scarcity months* and *conflicts in 2014* are significant with positive coefficients: the higher these covariates the higher the conflicts in 2015.
Furthermore, the posterior of λ is concentrated around 0.9 suggesting that the conflicts in 2015 in neighboring cells are really meaningful and useful in order to model our target. This feature is also shown by the posterior predictive distribution that presents a coherent spatial structure with respect to the real observed data.
Therefore we can conclude that, to properly learn and predict the spatial structure observed in the real data, it is crucial to consider the spatial autocorrelation of the conflicts in 2015.

References

[1] Arab, A. (2015). Spatial and Spatio-Temporal Models for Modeling Epidemiological Data with Excess Zeros *International Journal of Environmental Research and Public Health*, **12**, page 10536-10548; doi:10.3390/ijerph120910536

[2] Galli, N. (2019). Natural resources availability as a boosting factor for human conflict in the Lake Chad Basin, Master Thesis, Politecnico di Milano

[3] Lambert, D., Brown, J.P. and Florax R. J.G.M.(2010). A two-step estimator for a spatial lag model of counts: Theory, small sample performance and an application, *Regional Science and Urban Economics*, **40**, 241–252

[4] Simões, M. Carvalho, Aleixo, Gomes and Natário (2017). A Spatial Econometric Analysis of the Calls to the Portuguese National Health Line, *Econometrics*, **5**, pp. doi:10.3390/econometrics5020024

[5] Joseph, M. (2016). Exact sparse CAR models in Stan, https://mc-stan.org/users/documentation/case-studies/mbjoseph-CARStan.html

[6] Jin, Bradley P. and Banerjee (2005). Generalized hierarchical multivariate CAR models for areal data. Biometrics 61.4: 950-961.