TMA4250 : Project 3 : Gaussian Markov Random Fields

Yawar Mahmood

April 2024

Problem 1: Simulation and Visualization

a)

We are given neighbourhood structure matrices for both these areas, which indicate which areas neighbour each other. This data is contained in the **Admin1Graph.txt** and **Admin2Graph.txt**. Denoting the neighbour matrices by N, if area i and j border, $N_{ij} = 1$, else $N_{ij} = 0$. The diagonal entries are 0. The Besag model takes this structure and relates it to spatial correlation, through a precision matrix Q, given as $Q = \tau R$.

R is the structure matrix, and is given as such:

- **Diagonal Entries:** $R_{ii} = -\sum_{i \neq j} N_{ij}$. This sum equates to the negative count of neighbours for area i.
- **Off-diagonal Entries:** $R_{ij} = N_{ij}$ for $i \neq j$.

In our case, the dimension of Q_1 is 37×37, reflecting the 37 admin1 areas. The dimension of Q_2 is 775×775 based on the same reasoning. This is also verified numerically.

The following table summarizes the numerical confirmations for the ranks of precision matrices and the proportions of zeros, for the different areas:

Parameter	**Admin1**	**Admin2**
Rank of Q	36	774
Proportion of Zeros	0.1526662	0.008792508

Table 1: Rank and Zero Proportion Analysis

The ranks of the matrices Q_1 and Q_2 indicate their non-full rank nature, complicating the use of standard inversion techniques applicable to multivariate Gaussian distributions. Admin2, with a significantly lower proportion of zeros compared to Admin1, exhibits a more complex neighborhood structure, influencing the computational approach to these areas.

The sparcity structure of the precision matrices is given in 1 and 2.

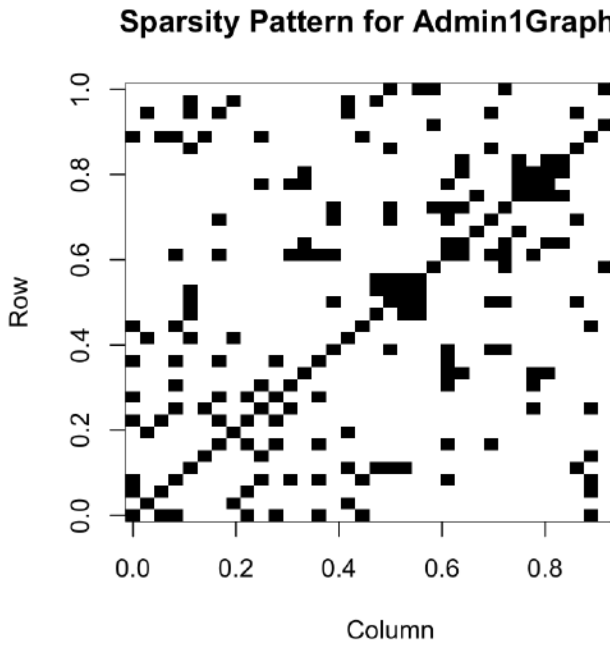

Figure 1: Sparsity pattern for Admin1

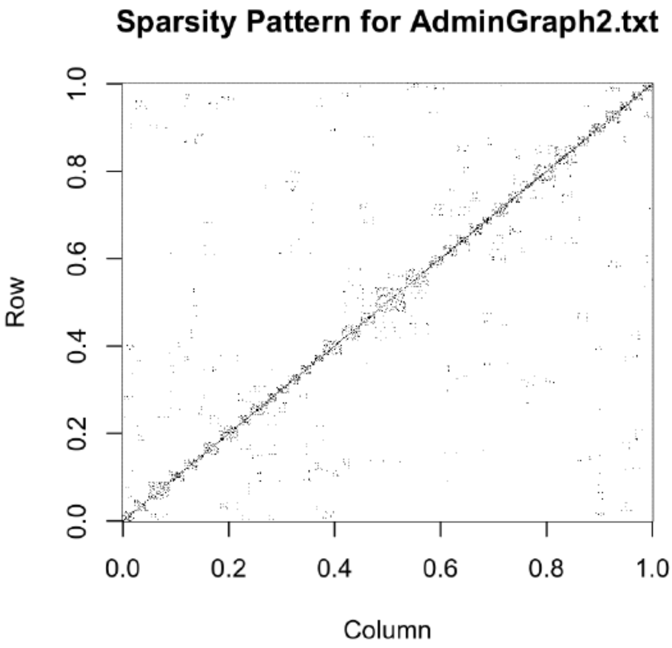

Figure 2: Sparsity pattern for Admin2

b)

We will now be simulating two GMRFs:

- The Besag model on the admin1 graph with $\tau_1 = 1$, and sum-to-zero constraint
- $N_{37}(0, I_{37})$

More about how to simulate, or draw samples, from these:

Simulating from the Besag Model on the Admin1 Graph

The implemented algorithm follows these steps to give out simulations:

1. **Construct the Precision Matrix (Q)**: This is done using the formula $Q = \tau * R$, where $\tau = 1$.
2. **Cholesky decomposition of Q**: To perform the decomposition, a small constant $1e-10$ is added to Q, to ensure numerical stability. Thereafter, a Cholesky decomposition is simply done by using the chol() function in R, obtaining L, such that $L^T L = Q$.
3. **Getting simulations**: Firstly, a standard normal matrix z is drawn from a $N(0, 1)$ distribution. Then, backsolving the system $L^T x = z$ for x, gives simulations from the normal distribution with mean 0 and covariance matrix Q^{-1} because $L^T L = Q$ and $LL^{-1} = I$

Simulating from $N_{37}(0, I_{37})$

This is simply done by using the mvrnorm() function in R

Sum-to-Zero Constraint

The sum-to-zero constraint ensures that the random fields values sum to zero across the different areas. This constraint is critical for identifiability. It essentially centers the field, making the interpretation of spatial effects relative rather than absolute, which makes the spatial model more interpretable.

Simulations

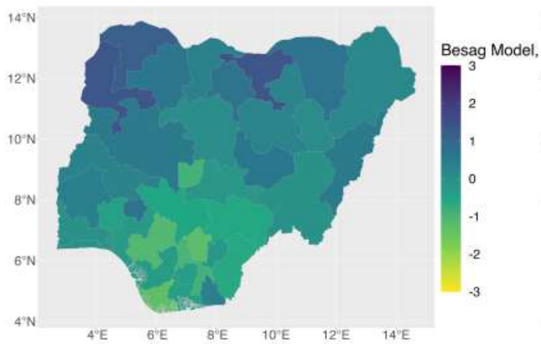

Figure 3: Realization one: Besag Model

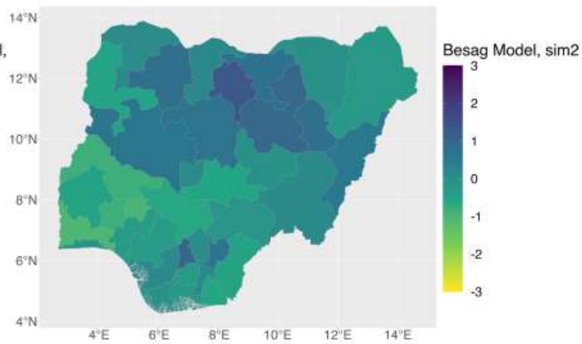

Figure 4: Realization two: Besag Model

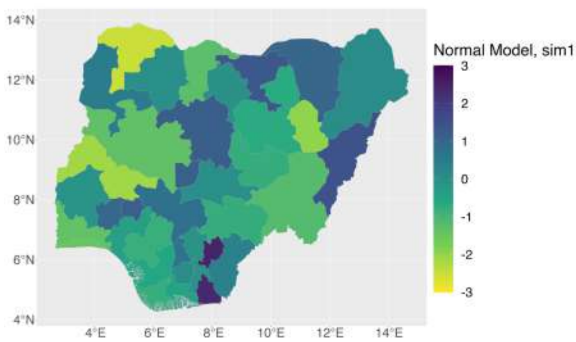

Figure 5: Realization one: Normal Model

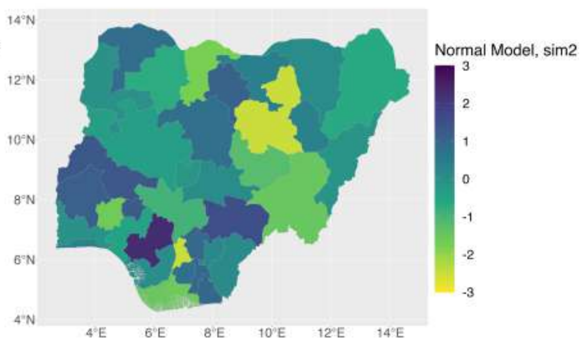

Figure 6: Realization two: Normal Model

Similarities and Differences

The differences between the Besag simulations and the Random Normal simulations are clear. The normal simulations seem more random, which makes sence, as the normal distribution has uniform variance across all the areas, with no specific dependence structure. The Besag model has a more stronger dependency structure, and as a result, neighbours tend to more similar, in contrast to the normal model. This results in that the Besag model produces more smoother spatial patterns.

c)

We will now run similar simulations, just using the admin2 graph instead:

- The Besag model on the admin2 graph with $\tau_1 = 1$, and sum-to-zero constraint
- $N_{775}(0, I_{775})$

Simulations

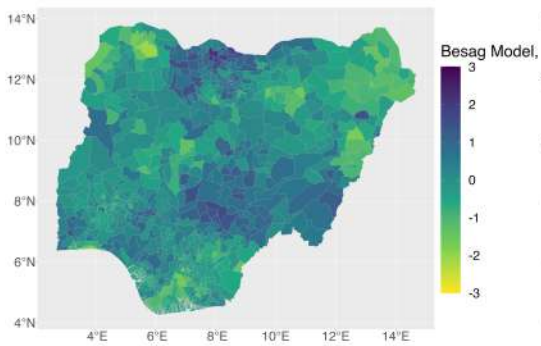

Figure 7: Realization one: Besag Model

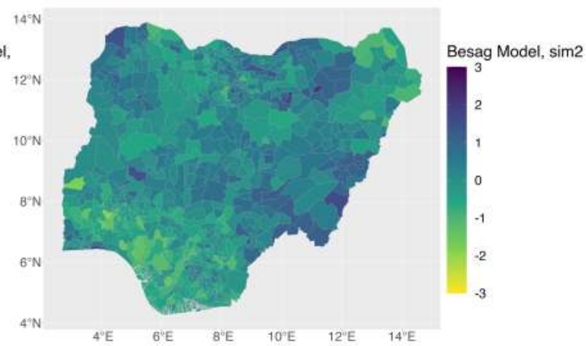

Figure 8: Realization two: Besag Model

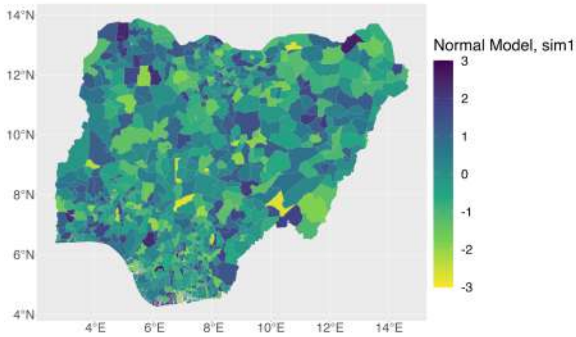

Figure 9: Realization one: Normal Model

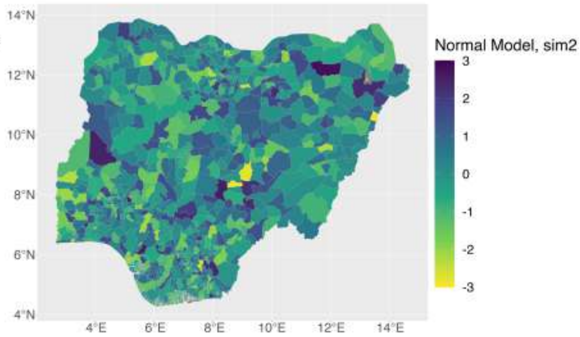

Figure 10: Realization two: Normal Model

Similarities and Differences

The same point raised in the previous discussion on the difference between the Besag simulations and the nornal simulations, based on the smoothness of the simulations and the dependency structure in the different cases, still apply.

The differences between the two distributions seems to be the same in the admin1 and admin2 graphs, just that admin2 is on a "bigger scale". That is, containing more neightbours.

d)

We now generate 100 realizations from the Besag model on the admin2 graph with $\tau_2 = 1$, with the sum-to-zero constraint.

We calculate the marginal variance in each admin2 area and display it on the admin2 map.

We also consider the area Gubio, and calculate the correlations between this admin2 area and all others, and display it on the admin2 map.

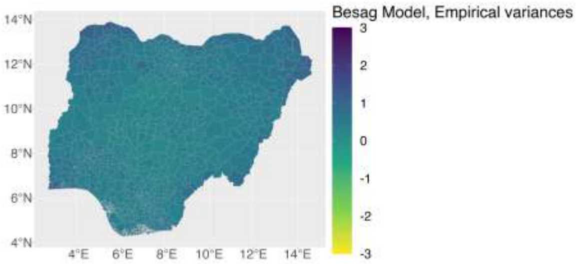

Figure 11: Marginal variance in each admin2 area

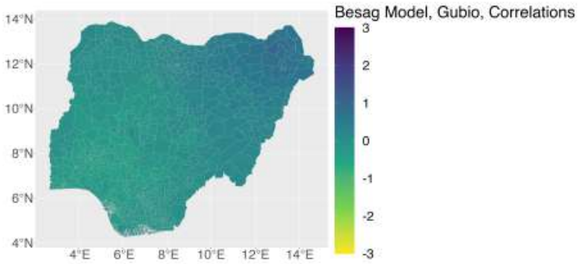

Figure 12: Correlations between Gubio and admin2 areas

The variance plot seems to be pretty constant, meaning that the variance values seems to be similar across the areas. There are some areas though, as in the top right corner of the map, where the variance values are bigger. And across the map there are small discrepancies. This indicates that the model is non-stationary. These discrepancies are very small, so I dont believe a stationary approximation would give the worst results on this model.

The correlation plot between Gubio, and the rest of the areas in admin2, seem to give an unexpected and interesting results. Gubio lies in the upper right corner of the map, and exhibits alot of correlation with its neightbours in that area. The interesting result is that it seems to be correlated on all other areas all over the country, even though the pairwise Markov Property indicates that non-neighbours are conditionally independent. The correlation here is a result of the Besag model. The Makrov property does not guarantee unconditional indepenence, meaning that a chain of states can affect each other. This is highligheted by the sum-to-zero constraint on the model, as the sum over all areas has to equal zero, meaning that some areas, even far apart, must have positive/negative correlations.

Problem 2: Small Area Estimation

a)

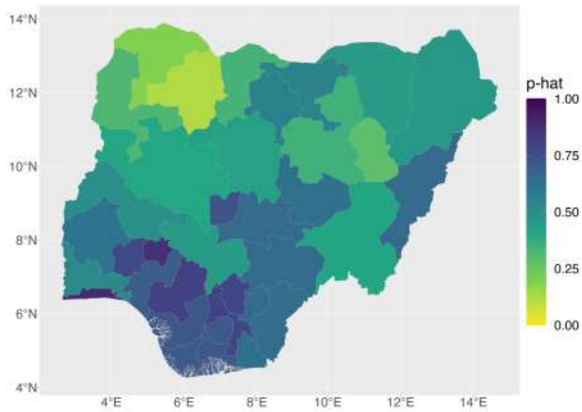

Figure 13: Observed proportions

The vaccination coverage data shows that the coverage is good at the north, and falls as one moves towards the south. Such a coverage trend, which is area based, hints at using a spatial model. This is only an assumption, and using a spatial mdoel may introduce bias to the model. This may be justified if the reduction in variance is significant.

b)

Finding the distribution: $Y \mid X$

Let p_a denote the true proportion of children who are vaccinated in area a. Let $\hat{P}_a$ be the estimator for p_a. A common assumption is:

$$\text{logit}(\hat{P}_a) \sim N(\text{logit}(p_a), V_a), \quad a = 1, \dots, 37,$$

Here, we use the notation that $X = (\text{logit}(P_1), \dots, \text{logit}(P_{37}))^T$ represents the random vector of true logit-transformed proportions, and $Y = (\text{logit}(\hat{P}_1), \dots, \text{logit}(\hat{P}_{37}))^T$ represents the random vector of observed logit-transformed proportions.

We want to find the distribution of $Y \mid X$. Given all this, we can simply deduce that:

$$Y \mid X \sim N(X, \Sigma),$$

where Σ is a diagonal matrix with diagonal entries $V_1, V_2, \dots, V_{37}$.

This implies that the observations $\text{logit}(\hat{P}_a)$ are conditionally independent, which is what we wanted. A key feature of the procedure is that there is no sharing of information between admin1 areas. That is, no spatial modelling.

Finding the distribution: $X \mid Y = y$

With a vague prior given by $X \sim N_{37}(0, \sigma^2 I_{37})$, where $\sigma^2 = 100^2$, we derive the conditional distribution of X given $Y = y$:

$$X \mid Y = y \sim N(\mu_{X|Y}, \Sigma_{X|Y}),$$

where

$$\mu_{X|Y} = \sigma^2(\sigma^2 I + \Sigma)^{-1} y, \quad \Sigma_{X|Y} = (\Sigma^{-1} + (\sigma^2 I)^{-1})^{-1}.$$

Given Σ is diagonal with entries $V_1, \ldots, V_{37}$, this simplifies to:

$$\Sigma_{X|Y} = \text{diag}\left(\frac{\sigma^2 V_a}{\sigma^2 + V_a}\right), \quad \mu_{X|Y} = \text{diag}\left(\frac{\sigma^2}{\sigma^2 + V_a}\right) y.$$

Asymptotic Behavior of $X \mid Y = y$

As $\sigma^2 \to \infty$, the prior's influence diminishes, leading to:

$$\mu_{X|Y} \approx y, \quad \Sigma_{X|Y} \approx \Sigma.$$

In this limit, $P_a \mid Y = y \sim N(y, \Sigma)$.

Median and coefficient of variation for $P_a \mid Y = y$

We now calculate the median and coefficient of variation for $P_a \mid Y = y$ based on 100 samples, and display them in two maps of the admin1 areas.

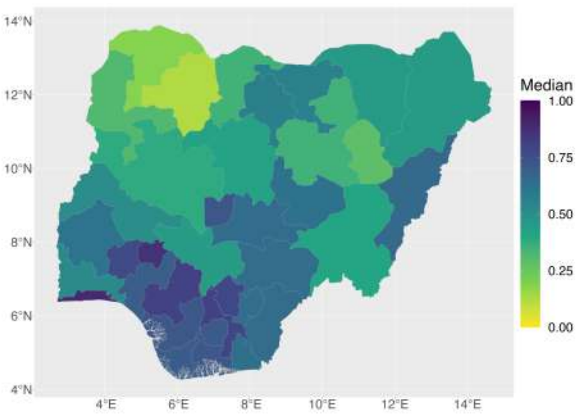

Figure 14: Median for $P_a \mid Y = y$

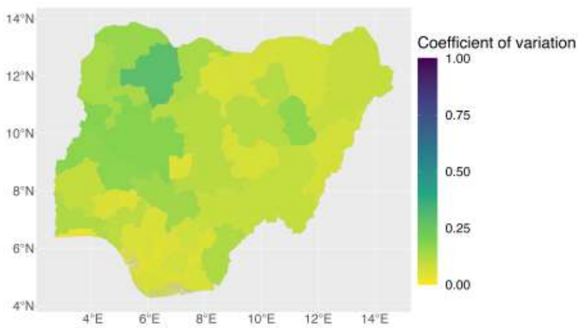

Figure 15: Coefficient of variation for $P_a \mid Y = y$

Comparing the results from the hierarchical spatial model, which is FIGURE X, to the actual observed proportions in FIGURE Y, we see that they match up pretty well. Aditionally, the variance is low, close to zero in many places. The choice of prior has has been good here, and it is to be expected, as we have not assumed any spatial dependency in this model, just as in the real case.

c)

We now assume that X a priori follows a Besag model. A Besag model is defined by:

$$Q = \tau(D - adj_{matrix})$$

Were adj_{matrix} is the neighbourhood matrix and D is a diagonal matrix where the entries are the number of neighbours for each area.

From a Baysian perspective, we then have:

$$Y|X \sim \mathcal{N}(X, V)$$
$$X \sim \mathcal{N}(0, Q^{-1})$$

Then the posterior distribution $X \mid Y = y$ is also multivariate normal. We can use the general formulas for the mean and the covariance of the posterior to find expression for these:

$$\Sigma_{\text{posterior}} = \left(\Sigma_0^{-1} + V^{-1}\right)^{-1}$$

$$\mu_{\text{posterior}} = \Sigma_{\text{posterior}} \left(\Sigma_0^{-1}\mu_0 + V^{-1}Y\right)$$

Where:

- $\mu_{\text{posterior}}$ is the mean of the posterior distribution.
- $\Sigma_{\text{posterior}}$ is the covariance matrix of the posterior distribution.
- Σ_0 is the covariance matrix of the prior distribution
- μ_0 is the mean of the prior distribution.
- V is the covariance matrix of the observational model
- Y is the observed data.

This gives, in our case:

$$X \mid Y = y \sim N((V^{-1} + Q)^{-1}(V^{-1}Y), (V^{-1} + Q)^{-1})$$

We now calculate the median and coefficient of variation for $P_a \mid Y = y$ based on 100 samples, and display them in two maps of the admin1 areas.

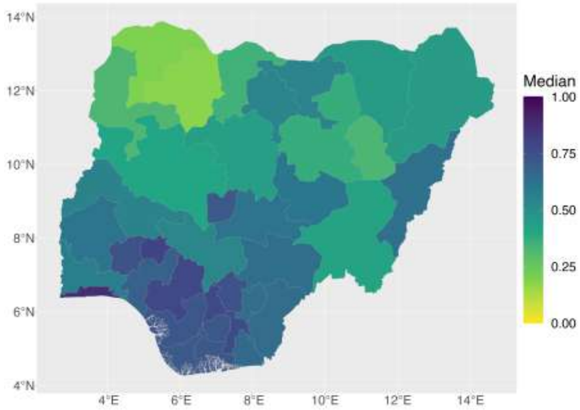

Figure 16: Median for $P_a \mid Y = y$

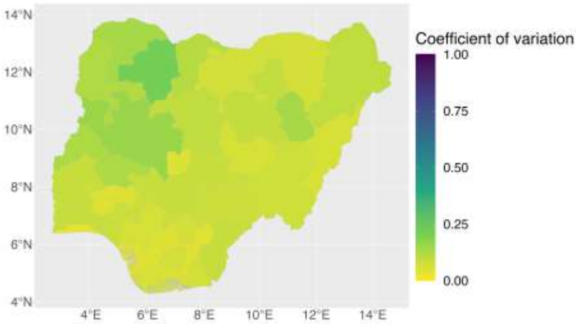

Figure 17: Coefficient of variation for $P_a \mid Y = y$

Even though the Besag simulations and the hierarchical model simulations gives similar results, there are some minor differences. The Besag model is generally smoother, and the coefficient variation is lower. This indicates that introducing spatial bias to the model helped in reducing the variance.

d)

From a bayesian standpoint, now that we have more information, we need to update our precision matrix and our mean.

The additional precision from the new survey in Kaduna, 0.1^{-2} is added to the precision matrix in the index of Kaduna.

The mean is updated to include the contribution from the new observation $y_{38} = 0.5$. This is added through the following expression:

$$\frac{1}{0.1^2}\text{logit}(0.5)$$

which expresses the additional precision-weighted logit-transformed observation from this area.

From here, the normal sample draws are done. We now calculate the median and coefficient of variation for $P_a \mid Y = y$ based on 100 samples, and display them in two maps of the admin1 areas.

Figure 18: Median for $P_a \mid Y = y$

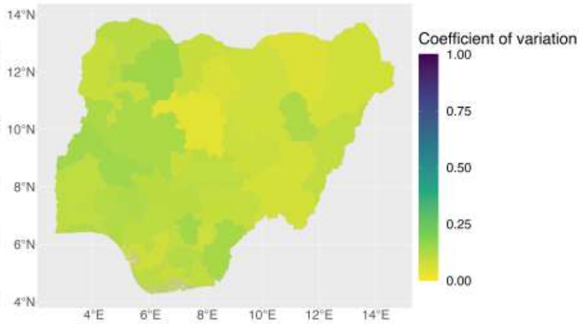

Figure 19: Coefficient of variation for $P_a \mid Y = y$

There seems to be something wrong with my implementation, as the medians just does not seem right. The variance is now lower that the information about Kaduna is known, which makes sence, as the uncertainty of Kaduna and their neighbours decreases.

e)

We run the same simulations, now for different values of τ.

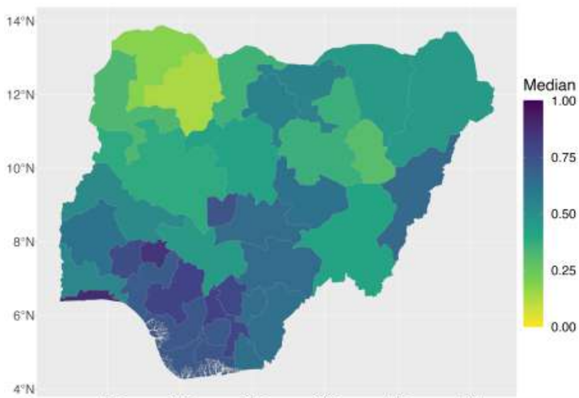

Figure 20: Median for $P_a \mid Y = y$, with $\tau = 0.1$

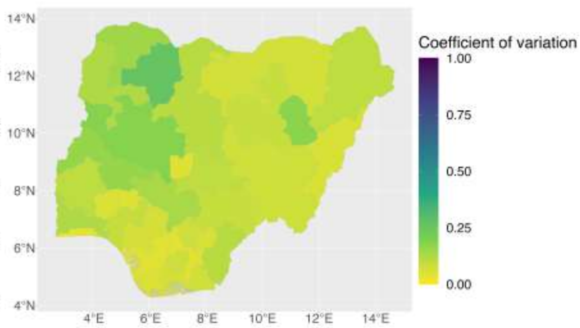

Figure 21: Coefficient of variation for $P_a \mid Y = y$, with $\tau = 0.1$

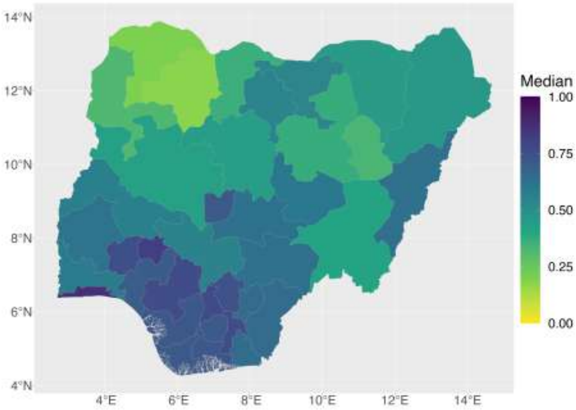

Figure 22: Median for $P_a \mid Y = y$, with $\tau = 1$

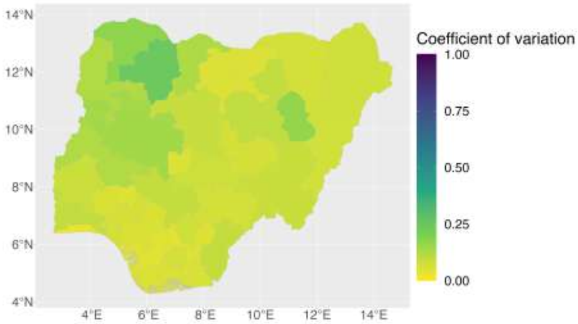

Figure 23: Coefficient of variation for $P_a \mid Y = y$, with $\tau = 1$

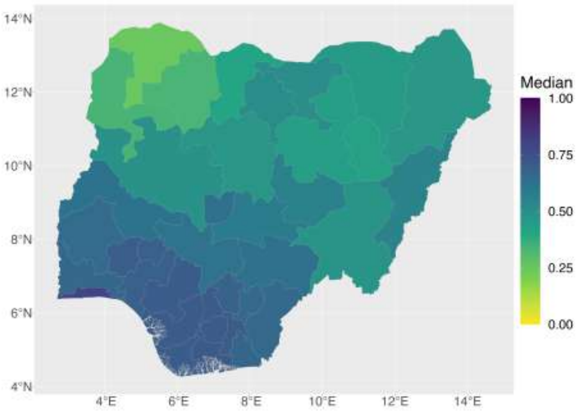

Figure 24: Median for $P_a \mid Y = y$, with $\tau = 10$

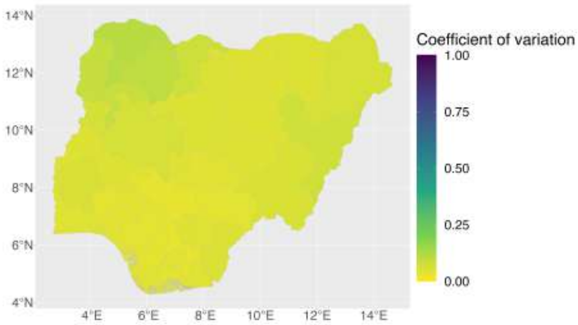

Figure 25: Coefficient of variation for $P_a \mid Y = y$, with $\tau = 10$

From the figures, we observe which affect different values of τ have on our simulations.

We see that for smaller values of τ, there is less spatial effect and makes the model less spatially dependent. Smaller τ does also increase the variance. These results makes sence, as τ is a parameter that controls the spatial relation. If the relation between areas is big, they are more likely the same, giving a smoother spatial structure and less variance.

f)

The optimizing routine was run 10 times and the average of these results is used as an estimator for τ. This came out to be: $\hat{\tau} = 6.279832$. The search area for τ was set to be between 0.1 and 10. The simulations are given below:

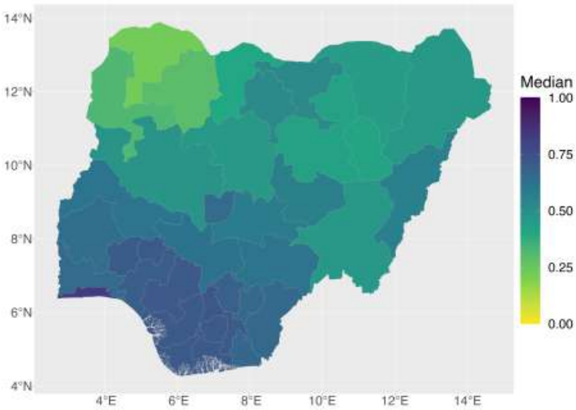

Figure 26: Median for $P_a \mid Y = y$, with optimal τ

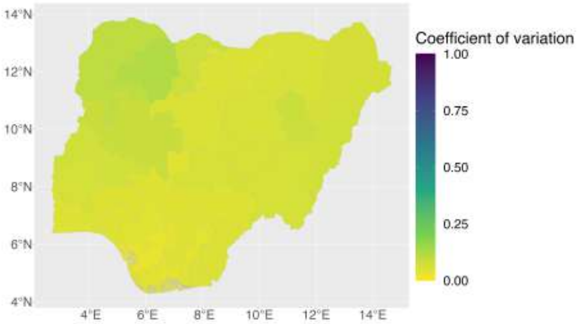

Figure 27: Coefficient of variation for $P_a \mid Y = y$, with optimal τ

We see that the results here are something in between the results for task e, for the values $\tau = 1$ and $\tau = 10$, as expected. The results are pretty similar to those obtained in task b) and c), but the results here have way less variarnce, and are way smoother. The smoothness is due to a higher spatial dependency as a result of a bigger τ value.