

TMA4250 Spatial Statistics

Project 3, Spring 2022

Christian Moen, Jim Totland

In this project we will study the geography and vaccination coverage of Nigeria, and use Gaussian Markov random fields (GMRFs) to model the relationships between regions on the map. The geography of Nigeria has two nested subdivisions which we will consider; *admin1*, which consists of 37 areas, and *admin2*, which consists of 775 areas.

Problem 1: Simulation and Visualization

a)

If we consider a Besag model \mathbf{X} with precision matrix $\mathbf{Q} = \tau\mathbf{R}$, we know that its density is given by

$$\begin{aligned} f_{\mathbf{X}}(\mathbf{x}; \tau) &\propto \exp\left\{-\frac{\tau}{2}\mathbf{x}^T\mathbf{R}\mathbf{x}\right\} \\ &= \exp\left\{-\frac{\tau}{2}\sum_{i \sim j}(x_i - x_j)^2\right\} \\ &= \exp\left\{-\frac{\tau}{2}\sum_{i \sim j}(x_i^2 + x_j^2) - \frac{\tau}{2}\sum_{i \sim j}(-2x_i x_j)\right\}, \quad \mathbf{x} \in \mathbb{R}^n \end{aligned}$$

from which we observe that the following must hold.

$$Q_{i,j} = \tau \begin{cases} |\text{ne}(i)|, & i = j \\ -1, & i \sim j \\ 0, & \text{else.} \end{cases} \quad (1)$$

If we represent the neighborhood structure as a matrix \mathbf{N} , where

$$N_{i,j} = \begin{cases} 1, & i \sim j \\ 0, & \text{else,} \end{cases}$$

we can express the precision matrix as

$$\mathbf{Q} = \tau(\text{diag}(\mathbf{N}\mathbf{1}) - \mathbf{N}),$$

where $\mathbf{1} = (1, \dots, 1)^T$. With this in mind, it is easy to construct the precision matrices of *admin1*, denoted $\mathbf{Q}_1 = \tau_1 \mathbf{R}_1$, and of *admin2*, denoted $\mathbf{Q}_2 = \tau_2 \mathbf{R}_2$. \mathbf{Q}_1 will be a 37×37 matrix, while \mathbf{Q}_2 will be a 775×775 matrix. Since the Besag model is a first-order intrinsic GMRF, the rank of \mathbf{Q}_1 will be $37 - 1 = 36$, while the rank of \mathbf{Q}_2 will be $775 - 1 = 774$.

The proportion of non-zero elements in \mathbf{Q}_1 is approximately 0.0270, while for \mathbf{Q}_2 it is approximately 0.00129. The sparsity patterns for each of the precision matrices is displayed in Figure 1.

Finally, we note that Besag model should be treated as a GMRF, and not as standard multivariate Gaussian distribution. This is because the precision matrix, \mathbf{Q} , is singular, and hence $|\mathbf{Q}| = 0$, which would make the probability density function of the multivariate Gaussian distribution undefined (the covariance

matrix does not exist). Furthermore, we want to preserve the sparsity in the model, which makes a GMRF with precision matrix \mathbf{Q} a natural choice.

b)

Now, we want to simulate from two GMRFs on the admin1 graph. First, we will simulate the Besag model with $\tau_1 = 1$ and a sum-to-zero constraint. This is done in line with Algorithm 1. The addition of $\varepsilon\mathbf{I}_n$ introduces an error in the eigenvalues which belong to \mathbf{Q} , but this error should be small, and the operation enables us to sample from a multivariate Gaussian distribution. We have chosen $\varepsilon = 10^{-10}$ throughout this section.

Algorithm 1 Simulation of first-order intrinsic GMRF ($\boldsymbol{\mu} = 0$).

Input: \mathbf{Q} : $n \times n$ precision matrix s.t. $\mathbf{Q}\mathbf{1} = 0$,

ε : small number.

1. $\tilde{\mathbf{Q}} \leftarrow \mathbf{Q} + \varepsilon\mathbf{I}_n$
 2. Compute Cholesky factor $\tilde{\mathbf{L}}$
 3. Sample $\mathbf{z} \sim \mathcal{N}_n(\mathbf{0}, \mathbf{I}_n)$
 4. Solve $\tilde{\mathbf{L}}^T \mathbf{v} = \mathbf{z}$
 5. $\mathbf{x} \leftarrow \mathbf{v} - \text{mean}(\mathbf{v})\mathbf{1}$
 6. **return** \mathbf{x}
-

We generate two realizations of the Besag model, which are displayed in Figure 2.

Next, we will simulate from $\mathcal{N}(\mathbf{0}, \mathbf{I}_{37})$, which is simply done with the function `rnorm` in R. We generate two realizations which are displayed in Figure 3.

Comparing the realizations in Figure 2 to the realizations in Figure 3, we observe that the realizations of the Besag model are less variable and neighboring regions tend to have more similar values, compared to the standard multivariate normal distribution. This is entirely in line with the Besag model, whose probability density function is constructed such that dissimilar neighbors are less probable by the factor $\sum_{i \sim j} -(x_i - x_j)^2$ in the exponential.

The sum-to-zero constraint means that the mean of the model is zero. In a way, we only model $\mathbf{Z} = \mathbf{X} - \bar{\mathbf{X}}$. This ensures identifiability of the model.

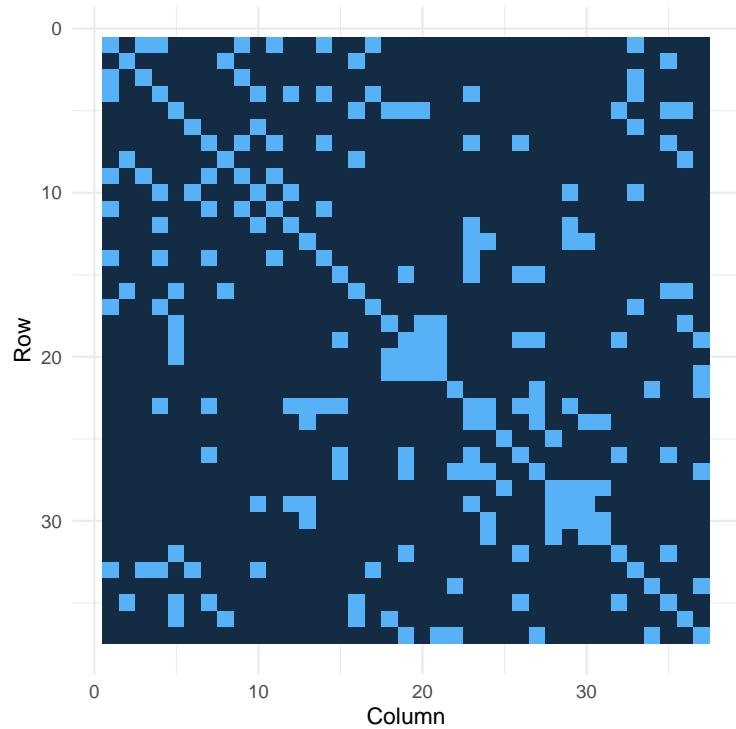
c)

Here, we will repeat the procedure from b), but on the admin2 graph. That is, we will generate two realizations from the Besag model with $\tau_2 = 1$ and a sum-to-zero constraint, and we will generate two realizations of $\mathcal{N}_{755}(\mathbf{0}, \mathbf{I}_{775})$, all on the admin2 graph. The realizations of the Besag model are displayed in Figure 4, while the realizations of the multivariate Gaussian are displayed in Figure 5.

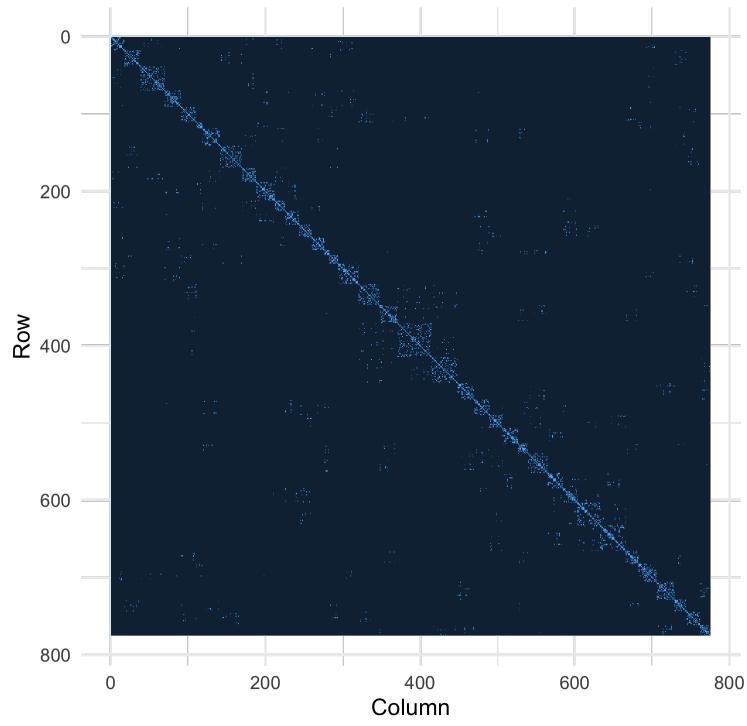
We observe that the multivariate Gaussian has higher variability than that of the Besag model, and the Besag model gives rise to more similar neighbors. This is expected, as discussed in b). The difference between the two GMRFs is perhaps more clearly illustrated here, on the admin2 graph, since it more clearly illustrates that the Besag models has far fewer 'extreme' values (if any) compared to the multivariate Gaussian. Additionally, there are more neighboring regions on the admin2 graph, which makes it more evident that there are 'areas' of similar regions and very few abrupt or sharp transitions between regions.

d)

We generate 100 realization of the Besag model described in c) on the admin2 graph, and compute the marginal variance for each domain, empirically. This is displayed in Figure 6. From the display, we observe that the empirical variance is different from region to region, which indicates that the Besag model is non-stationary in the sense that the marginal variance depends on the region. Furthermore, it would be reasonable if regions with more neighbors have lower marginal variance by the definition of the precision matrix in (1). This seems to be the case from the display, where we observe that the most variable regions have only 1-2 neighbors.



(a)



(b)

Figure 1: The sparsity patterns of (a) \mathbf{Q}_1 and (b) \mathbf{Q}_2 . The lighter blue indicates non-zero elements.

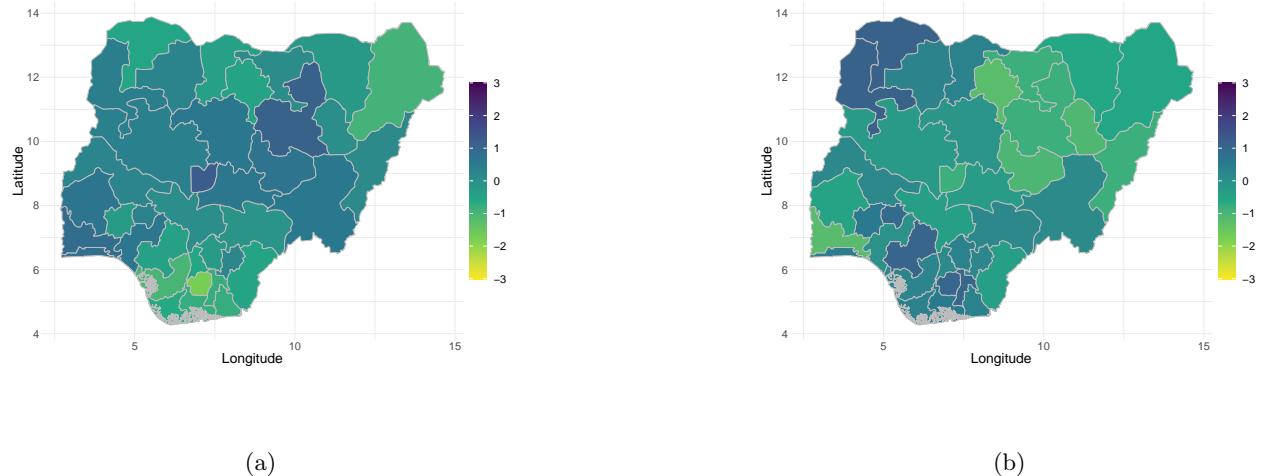


Figure 2: Two realizations, (a) and (b) of the Besag model with $\tau_1 = 1$ and a sum-to-zero constraint on the Nigeria admin1 graph.

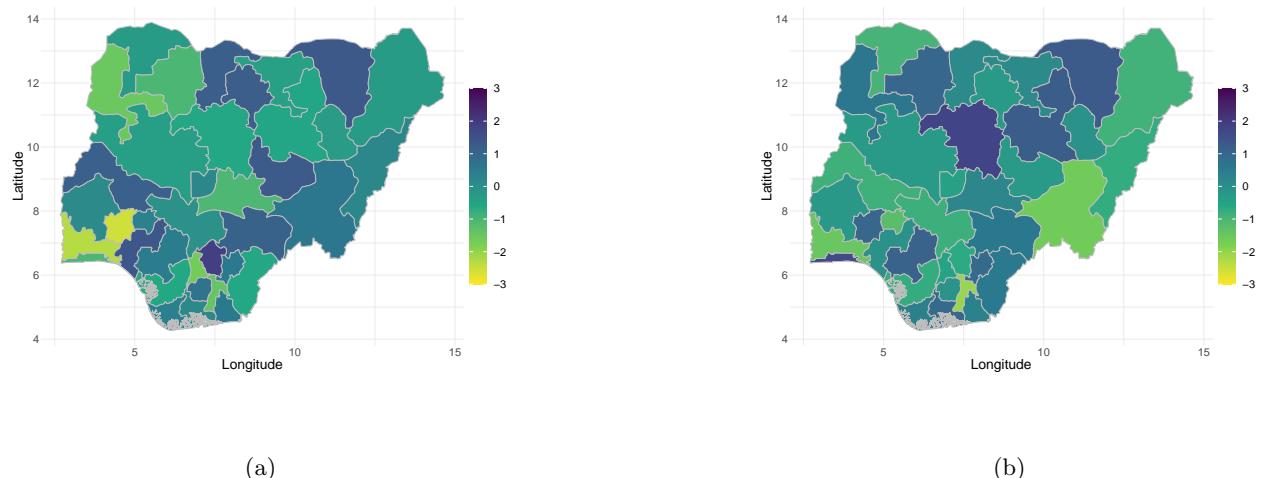


Figure 3: Two realizations, (a) and (b) of $\mathcal{N}(\mathbf{0}, \mathbf{I}_{37})$ on the Nigeria admin1 graph.

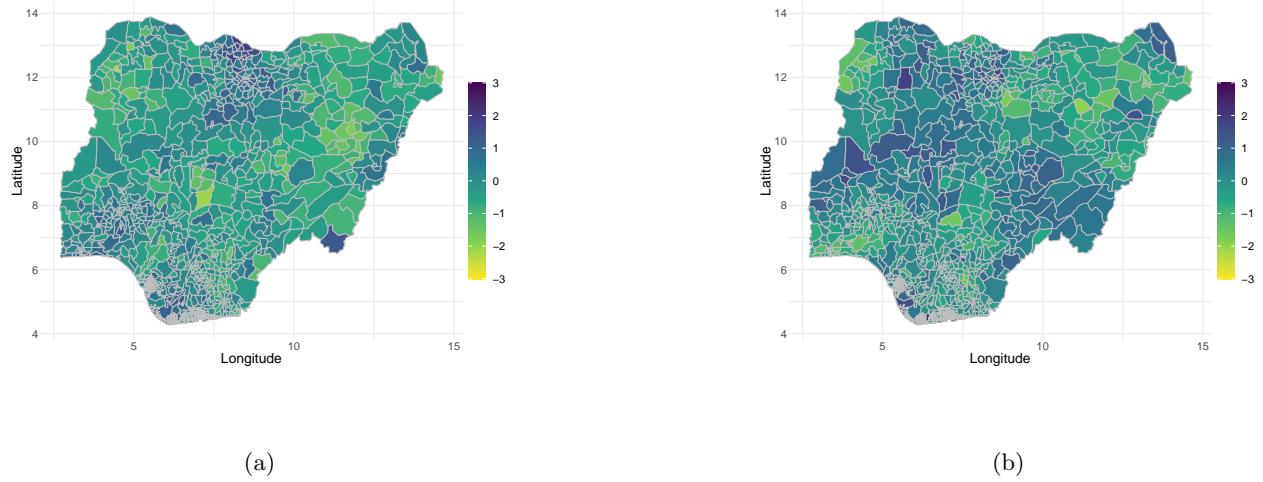


Figure 4: Two realizations, (a) and (b) of the Besag model with $\tau_2 = 1$ and a sum-to-zero constraint on the Nigeria admin2 graph.

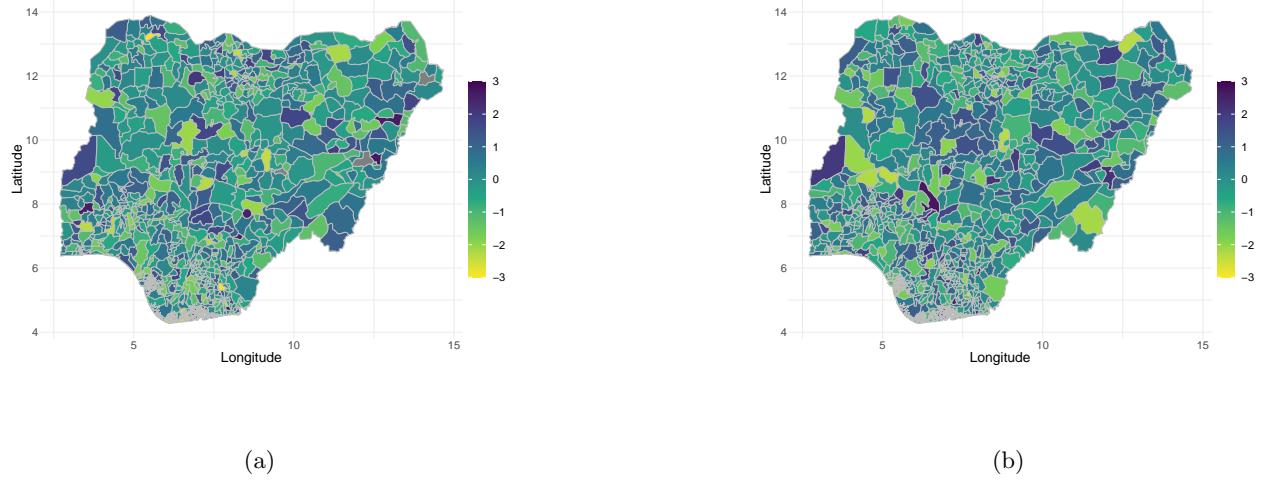


Figure 5: Two realizations, (a) and (b) of $\mathcal{N}_{775}(\mathbf{0}, \mathbf{I}_{775})$ on the Nigeria admin2 graph.

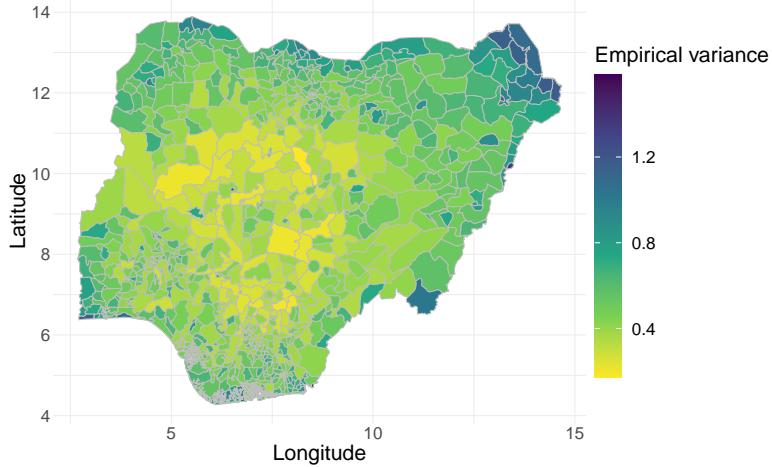


Figure 6: The empirical marginal variance of the regions on the admin2 graph, based on 100 realizations from the Besag model described in b).

Next, we specifically consider the region Gubio, which corresponds to admin2 area 150. Based on the same 100 realizations, we compute the empirical correlation between this admin2 area and all the others. This is displayed in Figure 7. From the display, we observe that there are both positive and negative correlations, and that it appears like a gradient over the map, where the correlation decreases with distance. Even though the Besag model satisfies the pairwise Markov property, i.e. $X_i \perp\!\!\!\perp X_j | \mathbf{X}_{-\{i,j\}}$, areas far apart can still be dependent (as the display illustrates). Informally, we can explain this by the fact that information can 'flow' between nodes (regions) when there is a path connecting them. For example, let us consider not an immediate neighbor of Gubio, but a neighbor of one of its immediate neighbors, say X_k . If we condition on X_k , this will alter the distribution of (at least) one of Gubio's neighbors. This, again, will alter the distribution of Gubio. Hence, Gubio is not unconditionally independent of X_k , even though they are not neighbors.

To explain the negative correlations in Figure 7, we refer to the sum-to-zero constraint. Since the immediate neighbors of Gubio will be similar and hence positively correlated, other regions must serve as a 'counterweights' to fulfill the sum-to-zero constraint, which gives rise to the negative correlations. Consequently, we can consider the negative correlations as an artifact of the sum-to-zero constraint.

Problem 2: Small Area Estimation

Nigeria does not have a complete registration system for vaccines which make it challenging to estimate *vaccination coverages* for each of the admin1 areas. There has been a huge effort to collect vaccination data which has made it possible to make a direct estimation of the vaccination proportion, denoted $\hat{\mathbf{P}}_a$, for each of the admin1 areas, denoted $a = 1, \dots, 37$.

We will apply hierarchical spatial models to the observations where we consider the true proportions to be stochastic variables. We use the notation $\mathbf{X} = (\text{logit}(P_1), \dots, \text{logit}(P_{37}))^\top = \text{logit}(\mathbf{P})$ and $\mathbf{Y} = (\text{logit}(\hat{P}_1), \dots, \text{logit}(\hat{P}_{37}))^\top = \text{logit}(\hat{\mathbf{P}})$ for the random vector of true proportions and the random vector of observed proportions, respectively. For such procedures mentioned above, a common assumption is that $\mathbf{Y} \sim \mathcal{N}(\text{logit}(p_a), V_a)$, where $\text{logit}(x) = \log(x/(1-x))$ and $\mathbf{V} = (V_1, \dots, V_{37})^\top$ are known variances.

a)

In Figure 8, the observed vaccination proportion, $\hat{\mathbf{p}}$, is plotted for each area in admin1. The display shows some spatial structure; the areas in the north appear to have a lower proportion of the population vaccinated than the southern areas, and neighboring areas seem to be correlated. This suggests that borrowing strength in space is a reasonable approach to reducing uncertainty.

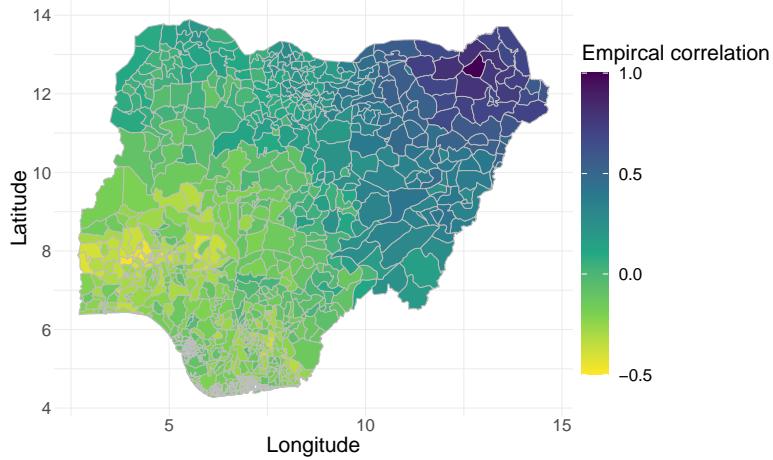


Figure 7: The empirical correlations between the area Gubio and all other areas on the admin2 graph, based on 100 realizations from the Besag model described in b).

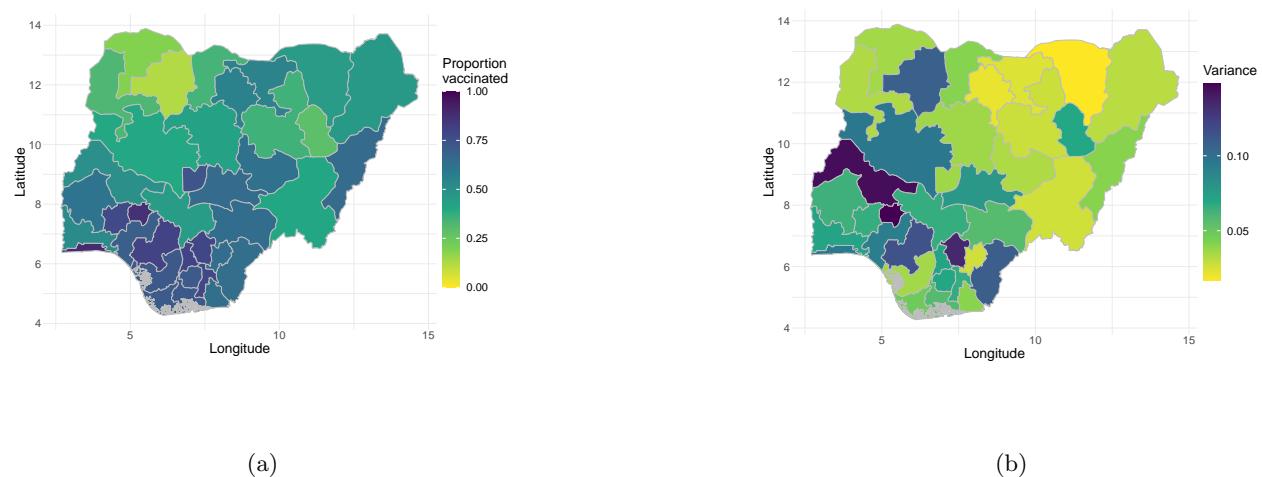


Figure 8: The direct estimation vaccination coverage (a) and its variance (b) of each admin1 region.

b)

To use the information in our observations \mathbf{Y} to further investigate the true vaccination proportion \mathbf{X} , we intend to use the posterior GMRF $\mathbf{X}|\mathbf{Y}$. If we assume a multivariate normal prior on \mathbf{X} and define $\mathbf{Z} = [\mathbf{X}, \mathbf{Y}]$ with mean $\boldsymbol{\mu} = [\boldsymbol{\mu}_{\mathbf{X}}, \boldsymbol{\mu}_{\mathbf{Y}}]^T$ and precision matrix

$$\mathbf{Q} = \begin{bmatrix} \mathbf{Q}_{xx} & \mathbf{Q}_{xy} \\ \mathbf{Q}_{yx} & \mathbf{Q}_{yy} \end{bmatrix},$$

we can write

$$\begin{aligned} \mathbf{X}|\mathbf{Y} &\sim \mathcal{N}(\boldsymbol{\mu}_{\mathbf{X}|\mathbf{Y}}, \mathbf{Q}_{\mathbf{X}|\mathbf{Y}}^{-1}), \text{ with} \\ \boldsymbol{\mu}_{\mathbf{X}|\mathbf{Y}} &= \boldsymbol{\mu}_{\mathbf{X}} - \mathbf{Q}_{xx}^{-1} \mathbf{Q}_{xy} (\mathbf{y} - \boldsymbol{\mu}_{\mathbf{Y}}) \text{ and} \\ \mathbf{Q}_{\mathbf{X}|\mathbf{Y}} &= \mathbf{Q}_{xx}. \end{aligned} \quad (2)$$

Let $f_{\mathbf{X}|\mathbf{Y}}$, $f_{\mathbf{Y}|\mathbf{X}}$ and $f_{\mathbf{Y}}$ denote the densities of $\mathbf{X}|\mathbf{Y}$, $\mathbf{Y}|\mathbf{X}$ and \mathbf{Y} respectively. Then, the joint density of \mathbf{Z} can be written as

$$f_{\mathbf{Z}}(\mathbf{x}, \mathbf{y}) = f_{\mathbf{X}}(\mathbf{x}) f_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x}), \quad (3)$$

where the conditional

$$\mathbf{Y}|\mathbf{X} = \mathbf{Y}|\text{logit}(\mathbf{P}) = \mathbf{Y}|\mathbf{P} \sim \mathcal{N}(\text{logit}(\mathbf{P}), \mathbf{Q}_{\mathbf{Y}|\mathbf{X}}^{-1}). \quad (4)$$

Here, the variance is given by $\mathbf{Q}_{\mathbf{Y}|\mathbf{X}}^{-1} = \text{diag}(\mathbf{V})$ for the known variance vector \mathbf{V} . Then, by placing a vague normal prior on \mathbf{X} with mean $\boldsymbol{\mu}_{\mathbf{X}} = \mathbf{0}$ and precision matrix $\mathbf{Q}_{\mathbf{X}} = \sigma^{-2}\mathbf{I}$, the joint density is

$$\begin{aligned} f_{\mathbf{Z}}(\mathbf{x}, \mathbf{y}) &\propto \exp \left\{ -\frac{1}{2} \mathbf{x}^T \mathbf{Q}_{\mathbf{X}} \mathbf{x} \right\} \exp \left\{ -\frac{1}{2} (\mathbf{y} - \text{logit}(\mathbf{P}))^T \mathbf{Q}_{\mathbf{Y}|\mathbf{X}} (\mathbf{y} - \text{logit}(\mathbf{P})) \right\} \\ &= \exp \left\{ -\frac{1}{2} \mathbf{x}^T (\mathbf{Q}_{\mathbf{X}} + \mathbf{Q}_{\mathbf{Y}|\mathbf{X}}) \mathbf{x} + \mathbf{y}^T \mathbf{Q}_{\mathbf{Y}|\mathbf{X}} \mathbf{x} - \frac{1}{2} \mathbf{y}^T \mathbf{Q}_{\mathbf{Y}|\mathbf{X}} \mathbf{y} \right\}, \end{aligned} \quad (5)$$

where we used $\mathbf{X} = \text{logit}(\mathbf{P})$ and simplified. By construction, we know that

$$\begin{aligned} f_{\mathbf{Z}}(\mathbf{x}, \mathbf{y}) &\propto \exp \left\{ -\frac{1}{2} \mathbf{z}^T \mathbf{Q}_{\mathbf{z}} \right\} \\ &= \exp \left\{ -\frac{1}{2} \mathbf{x}^T \mathbf{Q}_{xx} \mathbf{x} - \mathbf{y}^T \mathbf{Q}_{yx} \mathbf{x} - \frac{1}{2} \mathbf{y}^T \mathbf{Q}_{yy} \mathbf{y} \right\}. \end{aligned} \quad (6)$$

By matching Equation (5) and (6), we obtain

$$\begin{aligned} \mathbf{Q}_{xx} &= \mathbf{Q}_{\mathbf{X}} + \mathbf{Q}_{\mathbf{Y}|\mathbf{X}} = \sigma^{-2}\mathbf{I} + \text{diag}(\mathbf{V})^{-1} \\ \mathbf{Q}_{yy} &= \mathbf{Q}_{\mathbf{Y}|\mathbf{X}} = \text{diag}(\mathbf{V})^{-1} \\ \mathbf{Q}_{yx} &= -\mathbf{Q}_{\mathbf{Y}|\mathbf{X}} = -\text{diag}(\mathbf{V})^{-1} \\ \mathbf{Q}_{xy} &= \mathbf{Q}_{yx}^T = -\text{diag}(\mathbf{V})^{-1}, \end{aligned}$$

Inserting this into (2), we see that the multivariate normal distribution of the conditional $\mathbf{X}|\mathbf{Y}$ has parameters

$$\begin{aligned} \boldsymbol{\mu}_{\mathbf{X}|\mathbf{Y}} &= -\mathbf{Q}_{xx}^{-1} \mathbf{Q}_{xy} \mathbf{Y} \\ &= -(\mathbf{Q}_{\mathbf{X}} + \mathbf{Q}_{\mathbf{Y}|\mathbf{X}})^{-1} (-\mathbf{Q}_{\mathbf{Y}|\mathbf{X}} \mathbf{Y}) \\ &= (\sigma^{-2}\mathbf{I} + \text{diag}(\mathbf{V})^{-1})^{-1} \text{diag}(\mathbf{V})^{-1} \mathbf{Y} \\ &= (\sigma^{-2} \text{diag}(\mathbf{V}) + \mathbf{I})^{-1} \mathbf{Y}, \text{ and} \end{aligned} \quad (7)$$

$$\mathbf{Q}_{\mathbf{X}|\mathbf{Y}} = \mathbf{Q}_{xx} = \sigma^{-2}\mathbf{I} + \text{diag}(\mathbf{V})^{-1}. \quad (8)$$

If we let $\sigma^2 \rightarrow \infty$ we see that the conditional mean becomes $\boldsymbol{\mu}_{\mathbf{X}|\mathbf{Y}} = \mathbf{Y}$ and the precision matrix becomes $\mathbf{Q}_{\mathbf{X}|\mathbf{Y}} = \text{diag}(\mathbf{V})^{-1}$. This is reasonable, since the prior on \mathbf{X} will not contain any information, so the expected value of $\mathbf{X}|\mathbf{Y}$ is the given data, namely \mathbf{Y} , and the variance is the variance of \mathbf{Y} . In this

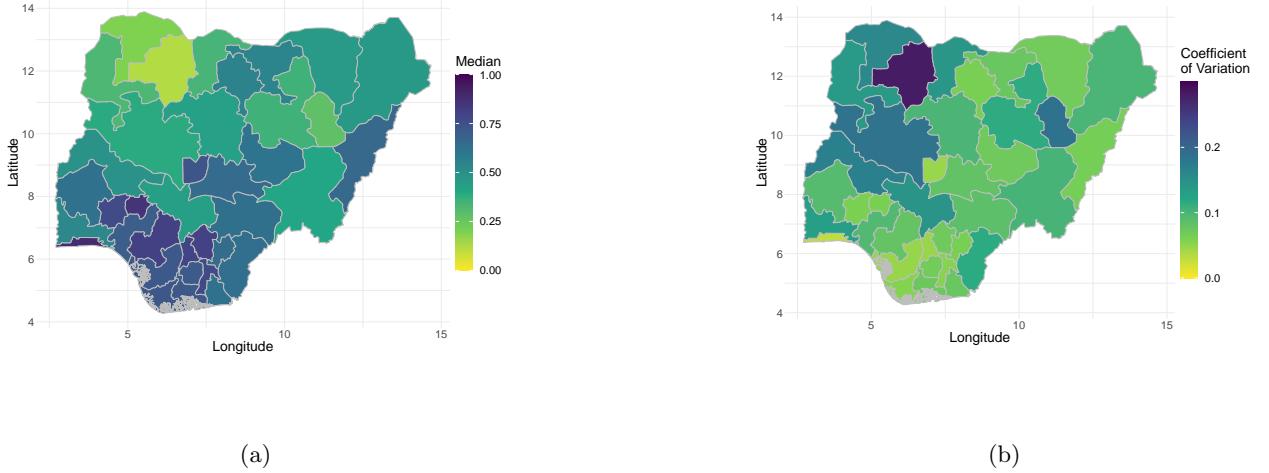


Figure 9: The median (a) and coefficient of variation (b) of 100 simulations of $P_a|\mathbf{Y}$, where \mathbf{X} is assumed to follow a normal distribution.

limiting case, $P_a|\mathbf{Y} = \mathbf{y} = \text{expit}(X_a)|\mathbf{Y} = \mathbf{y}$ will be logit-normally distributed with parameters $\mu_a = y_a$ and $\sigma_a^2 = V_a$ for $a = 1, \dots, 37$.

We generate 100 samples of $P_a|\mathbf{Y} = \mathbf{y}$ and display the median and coefficient of variation for each admin1 region in Figure 9. We see that the median looks very similar to the direct estimation vaccination proportion in Figure 8. The coefficient of variation and variance in these two figures show some differences. The range of the former is shifted higher and is also larger compared to the direct estimation variance in Figure 8. It is also smoother, that is, neighboring areas show more similarities than the direct estimation variation, while keeping the general trend of the coefficient of variation similar to the variance.

c)

We now assume that \mathbf{X} *a priori* follows a Besag model on the admin1 graph, i.e. \mathbf{X} is a first order intrinsic GMRF with precision matrix $\mathbf{Q}_\mathbf{X} = \tau \mathbf{R}_1$. Then, to find the expected value and precision matrix of $\mathbf{X} | \mathbf{Y} = \mathbf{y}$, we follow the same procedure as in b), and obtain

$$\begin{aligned}\mathbf{Q}_{xx} &= \mathbf{Q}_\mathbf{X} + \mathbf{Q}_{\mathbf{Y}|\mathbf{X}} = \tau \mathbf{R}_1 + \text{diag}(\mathbf{V})^{-1} \\ \mathbf{Q}_{yy} &= \mathbf{Q}_{\mathbf{Y}|\mathbf{X}} = \text{diag}(\mathbf{V})^{-1} \\ \mathbf{Q}_{yx} &= -\mathbf{Q}_{\mathbf{Y}|\mathbf{X}} = -\text{diag}(\mathbf{V})^{-1} \\ \mathbf{Q}_{xy} &= \mathbf{Q}_{yx}^\top = -\text{diag}(\mathbf{V})^{-1}.\end{aligned}$$

Inserting this into (2), we obtain the expected value and the precision matrix:

$$\begin{aligned}\boldsymbol{\mu}_{\mathbf{X}|\mathbf{Y}} &= (\tau \mathbf{R}_1 + \text{diag}(\mathbf{V})^{-1})^{-1} \text{diag}(\mathbf{V})^{-1} \mathbf{y}, \\ \mathbf{Q}_{\mathbf{X}|\mathbf{Y}} &= \tau \mathbf{R}_1 + \text{diag}(\mathbf{V})^{-1}.\end{aligned}\tag{9}$$

Assuming that all elements of \mathbf{V} are non-zero, we observe that (9) must have full rank, since we effectively remove the sum-to-zero constraint by adding $\text{diag}(\mathbf{V})^{-1}$, exactly as in Algorithm 1. Hence, $\mathbf{X} | \mathbf{Y} = \mathbf{y}$ is a proper GMRF.

Next, we assume $\tau = 1$, and generate 100 samples of $P_a|\mathbf{Y} = \mathbf{y}$, which is logit-normal and has parameters $\mu_a = (\boldsymbol{\mu}_{\mathbf{X}|\mathbf{Y}})_a$ and $\sigma_a^2 = (\mathbf{Q}_{\mathbf{X}|\mathbf{Y}})_{a,a}^{-1}$, for $a = 1, \dots, 37$. The median and coefficient of variation are displayed in Figure 10. Comparing to Figure 9, we observe that the pattern is the same for both plots, though a little

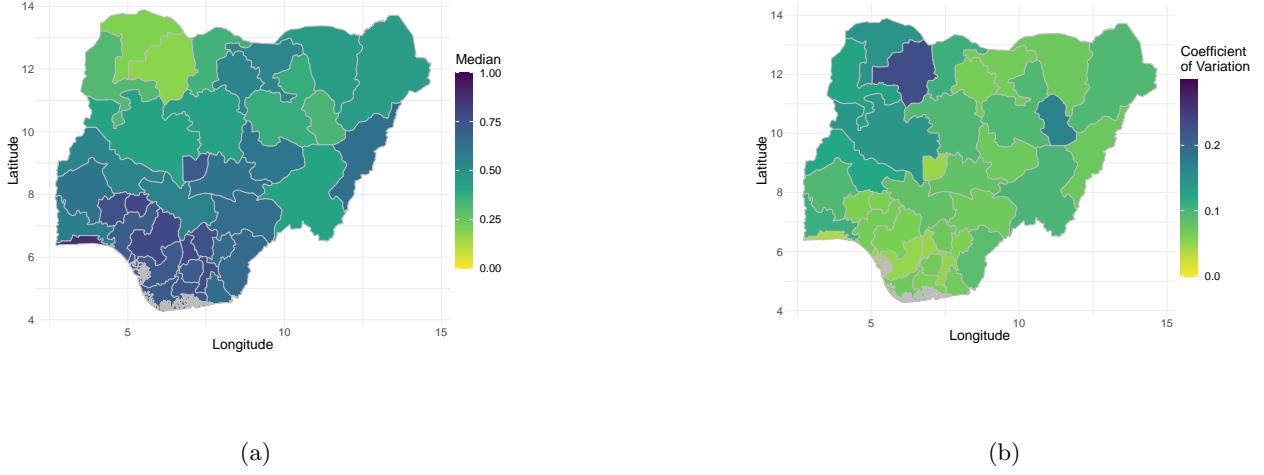


Figure 10: The median (a) and coefficient of variation (b) of 100 simulations of $P_a|\mathbf{Y}$, where \mathbf{X} is assumed to follow a Besag model and the precision parameter $\tau = 1$.

smoother, i.e. neighboring regions seem to be a little more similar with respect to both the median and the coefficient of variation. This is reasonable, since the Besag-prior should introduce a positive correlation between neighboring regions. Also, we observe that the coefficient of variation is lower for the Besag-model, which is a key component of borrowing strength in space. This may not be so evident from the displays, but we calculate that the mean coefficient of variation has decreased from 0.097 to 0.084. The model's sensitivity to τ will be further investigated in section e).

d)

Now, we imagine that in addition to the survey which gave rise to the observations \mathbf{Y} , an independent survey in (the admin1 area) Kaduna gave rise to a much more precise estimate of the proportion for that state. We assume that

$$Y_{38}|P_{\text{Kaduna}} \sim \mathcal{N}(\text{logit}(P_{\text{Kaduna}}), 0.1^2),$$

where $Y_{38}|P_{\text{Kaduna}}$ is independent of $\mathbf{Y}|\mathbf{P}$. Additionally, let $\tilde{\mathbf{Y}} = (Y_1, \dots, Y_{37}, Y_{38})^\top$ and assume that \mathbf{X} follows the same Besag model as in c). Now, we wish to determine the expected value, $\boldsymbol{\mu}$, and the precision matrix, \mathbf{Q} , of $\mathbf{X}|\tilde{\mathbf{Y}} = \tilde{\mathbf{y}}$. Let also $\tilde{\mathbf{V}} = (\mathbf{V}, 0.1^2)^\top$ and \mathbf{M} be a 38×37 matrix with 1 on the diagonal and 0 elsewhere, except for the last row, where the 1 is positioned on the index of Kaduna. Then, we can write

$$\tilde{\mathbf{Y}}|\mathbf{X} \sim \mathcal{N}(\mathbf{MX}, \text{diag}(\tilde{\mathbf{V}})).$$

Let $\mathbf{Q}_1 = \tau \mathbf{R}_1$ be the precision matrix of the Besag model, and $\mathbf{Q}_2 = \text{diag}(\tilde{\mathbf{V}})^{-1}$. By following the same procedure as in b), we obtain the expected value and precision matrix given by

$$\begin{aligned} \boldsymbol{\mu} &= [\mathbf{Q}_1 + \mathbf{M}^\top \mathbf{Q}_2 \mathbf{M}]^{-1} \mathbf{M}^\top \mathbf{Q}_2 \tilde{\mathbf{Y}}, \\ \mathbf{Q} &= \mathbf{Q}_1 + \mathbf{M}^\top \mathbf{Q}_2 \mathbf{M}. \end{aligned}$$

Next, we assume that $y_{38} = 0.5$ and $\tau = 1$. We generate 100 samples of $P_a|\tilde{\mathbf{Y}} = \tilde{\mathbf{y}}$, which is logit-normally distributed with parameters $\mu_a = (\boldsymbol{\mu})_a$ and $\sigma_a^2 = (\mathbf{Q})_{a,a}^{-1}$, for $a = 1, \dots, 37$. We compute The median and coefficient of variation for the generated samples. The result is displayed in Figure 11. Comparing to Figures 9 and 10, we observe that the coefficient of variation is lower across almost the whole map, which is reasonable, since we add additional information to the posterior. This decrease in variation is clearest for

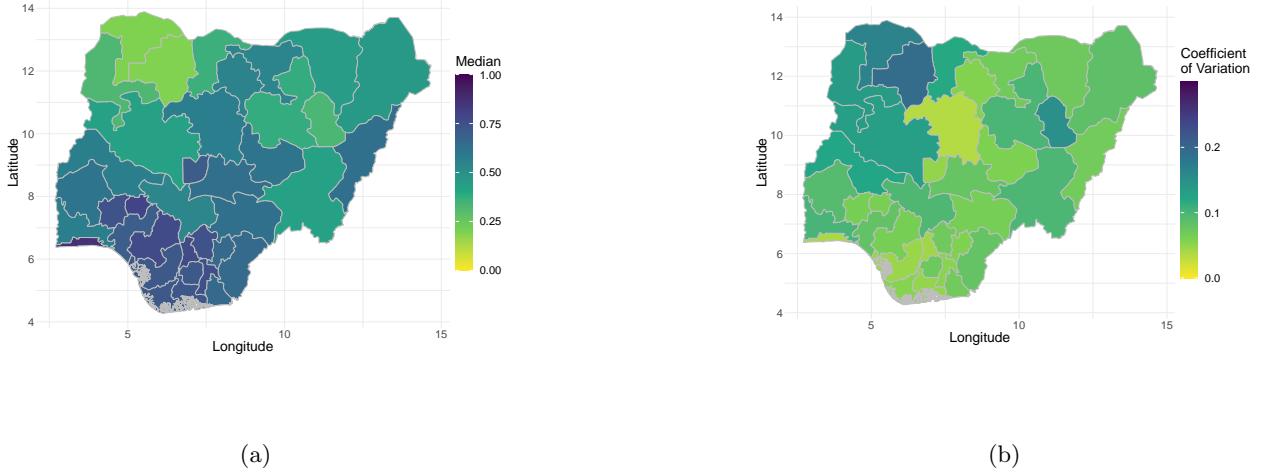


Figure 11: The median (a) and coefficient of variation (b) of 100 simulations of $P_a|\mathbf{X}$ with parameters $\mu_a = (\boldsymbol{\mu}_{\mathbf{X}|\mathbf{Y}})_a$ and $\sigma_a^2 = (\mathbf{Q}_{\mathbf{X}|\mathbf{Y}})_{a,a}^{-1}$, for $a = 1, \dots, 37$.

Kaduna (upper right middle). From the display of the median, we see that is quite similar to that in Figure 10, though we observe that the median for Kaduna increases slightly.

e)

In section c) we sampled $P_a|\mathbf{Y} = \mathbf{y}$, and \mathbf{X} was assumed to follow a Besag model with precision parameter $\tau = 1$. Now we will investigate the sensitivity of the model to τ by simulating 100 times for $\tau = 0.1$ and $\tau = 10$. In Figure 12 we see the median and coefficient of variation for the two values of τ . For $\tau = 0.1$, both the median and the coefficient of variation are more similar to the results from the samples where \mathbf{X} was assumed multivariate normal (section b)) and less smooth than when $\tau = 1$. When $\tau = 10$, the median and coefficient of variation become even smoother than when $\tau = 1$, and the differences from the results in section b) are greater. The Besag model borrows strength in space, that is, the precision parameter τ can be interpreted as similarity coefficient dictating how much the information in neighboring regions are used to predict the value of a region. This means that the similarities of the median of the proportion vaccinated between neighboring regions become greater. Also, we see that the coefficient of variability becomes more homogeneous and smaller in magnitude across areas. This can be attributed to the fact that neighboring regions are 'forced' to be more similar as τ increases, which leads to less variation. Thus, both the median and the coefficient of variation become more similar across regions as the precision parameter increases, and we would argue that a correct estimate of τ is important.

f)

In sections c) and e) we looked at some values for the precision parameter τ and saw that both the median and the coefficient of variation is sensitive to τ . Now we will try to find a better value for τ using the maximum likelihood estimate $\hat{\tau}$. Bayes rule gives

$$f(\mathbf{y}; \tau) = \frac{f(\mathbf{x}; \tau) f(\mathbf{y}|\mathbf{x})}{f(\mathbf{x}|\mathbf{y}; \tau)}.$$

Then, the log-likelihood is

$$\ell(\tau; \mathbf{y}) = \log(f(\mathbf{y}; \tau)) = \log(f(\mathbf{x}; \tau)) + \log(f(\mathbf{y}|\mathbf{x})) - \log(f(\mathbf{x}|\mathbf{y}; \tau)), \quad (10)$$

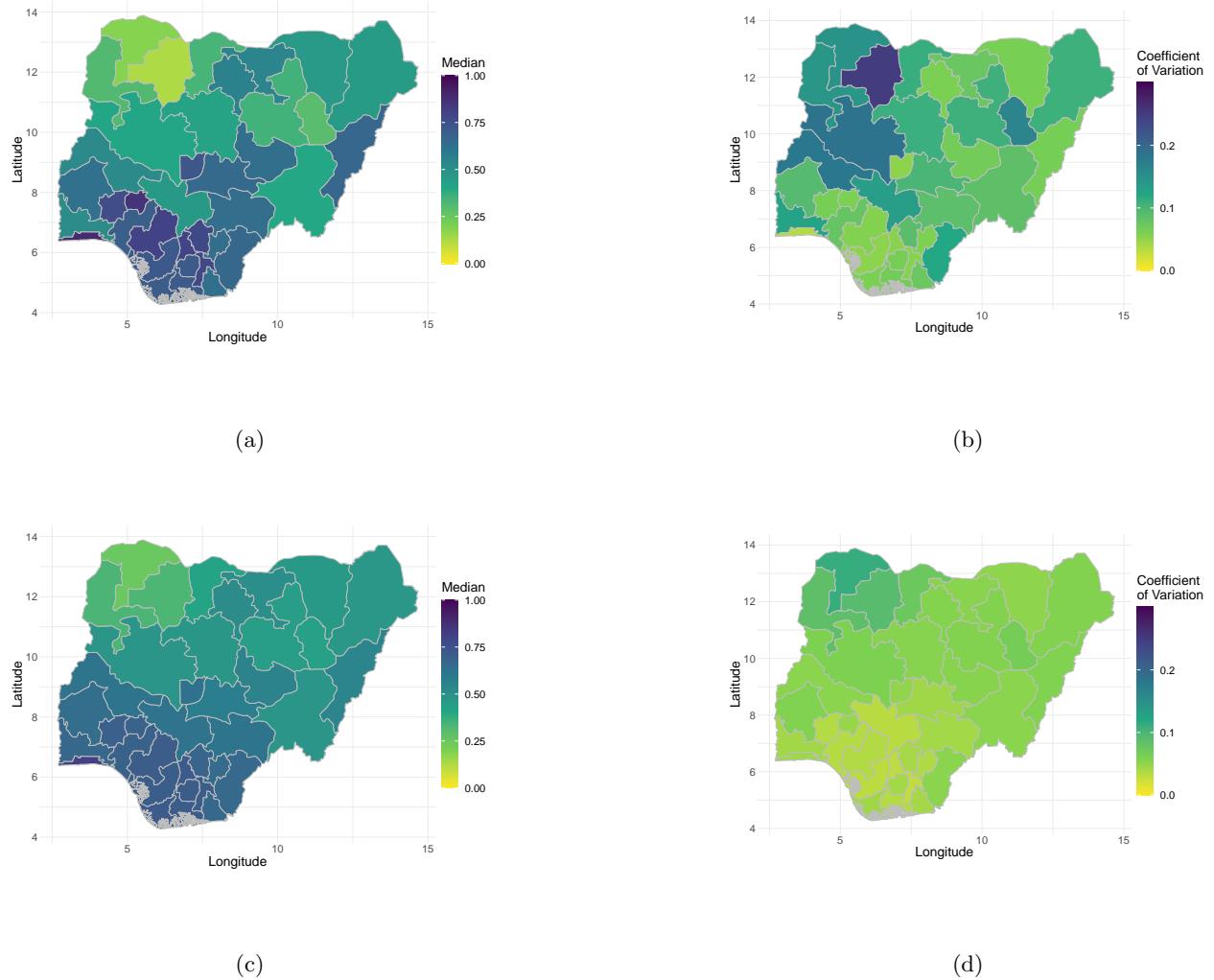


Figure 12: The median (a) and (c), and coefficient of variation (b) and (d) of 100 simulations of $P_a|\mathbf{Y}$, where \mathbf{X} is assumed to follow a Besag model and the precision parameter $\tau = 0.1$ and $\tau = 10$ for top and bottom row respectively.

where $\mathbf{X}|\mathbf{Y} \sim \mathcal{N}_{37}(\boldsymbol{\mu}_C, \mathbf{Q}_C^{-1})$, $\mathbf{Y}|\mathbf{X} \sim \mathcal{N}(\text{logit}(\mathbf{P}), \mathbf{D})$ and \mathbf{X} is a Besag model with structure matrix $\mathbf{R} := \mathbf{R}_1$. That is,

$$\begin{aligned} f(\mathbf{x}|\mathbf{y}; \tau) &= c_1 |\mathbf{Q}_C|^{1/2} \exp \left\{ -\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu}_C)^T \mathbf{Q}_C (\mathbf{x} - \boldsymbol{\mu}_C) \right\} \\ f(\mathbf{y}|\mathbf{x}) &= c_2 \exp \left\{ -\frac{1}{2} (\mathbf{y} - \mathbf{x})^T \mathbf{D}^{-1} (\mathbf{y} - \mathbf{x}) \right\} \\ f(\mathbf{x}; \tau) &= c_3 \tau^{\frac{37-1}{2}} \exp \left\{ -\frac{\tau}{2} \mathbf{x}^T \mathbf{R} \mathbf{x} \right\}. \end{aligned}$$

Then we insert the logarithm of the distributions into Equation (10) and obtain

$$\begin{aligned} \ell(\tau; \mathbf{y}) &= \text{Const} + \frac{37-1}{2} \log(\tau) - \frac{\tau}{2} \mathbf{x}^T \mathbf{R} \mathbf{x} - \frac{1}{2} (\mathbf{y} - \mathbf{x})^T \mathbf{D}^{-1} (\mathbf{y} - \mathbf{x}) \\ &\quad - \frac{1}{2} \log |\mathbf{Q}_C| + \frac{1}{2} (\mathbf{x} - \boldsymbol{\mu}_C)^T \mathbf{Q}_C (\mathbf{x} - \boldsymbol{\mu}_C). \end{aligned} \quad (11)$$

We want to maximize this quantity with respect to τ , that is we want to find

$$\hat{\tau} = \arg \max_{\tau} (\ell(\tau; \mathbf{y})).$$

This is done numerically in R using the function `optimize`, which yielded $\hat{\tau} \approx 0.806$.

We could consider the possibility of directly computing $\log(f(\mathbf{y}; \tau))$ using the marginal distribution of \mathbf{Y} . That is, we may compute

$$f(\mathbf{y}; \tau) = \int f(\mathbf{x}, \mathbf{y}; \tau) d\mathbf{x} = \int f(\mathbf{y}|\mathbf{x}) f(\mathbf{x}; \tau) d\mathbf{x}.$$

However, since \mathbf{X} has an improper distribution, we need to ensure that the resulting density is normalized. This is where we run into trouble, since \mathbf{Y} also has an improper density, meaning that we cannot normalize the resulting density. To see that this must be the case, we could consider the law of total variance:

$$\begin{aligned} \text{Var}(\mathbf{Y}) &= \text{E}[\text{Var}(\mathbf{Y}|\mathbf{X})] + \text{Var}[\text{E}(\mathbf{Y}|\mathbf{X})] \\ &= \text{diag}(\mathbf{V}) + \text{Var}(\mathbf{X}), \end{aligned}$$

which shows that the covariance matrix of \mathbf{Y} is undefined. Consequently, it would not be possible to calculate $\log(f(\mathbf{y}; \tau))$ 'safely' with this procedure.

Next, we generate 100 samples of $P_a|\mathbf{Y} = \mathbf{y}$ with the same parameters as in section c), μ_a and σ_a^2 , using the maximum likelihood estimate of the precision parameter, $\tau = \hat{\tau}$. In Figure 13, we observe that both the median and the coefficient of variation are smoother than their counterparts in Figure 9, but not as smooth as those in Figure 10. Additionally, the mean coefficient of variation is 0.087, which is lower than in b), but slightly larger than in c). These results coincide with the explanation of the precision parameter in section e).

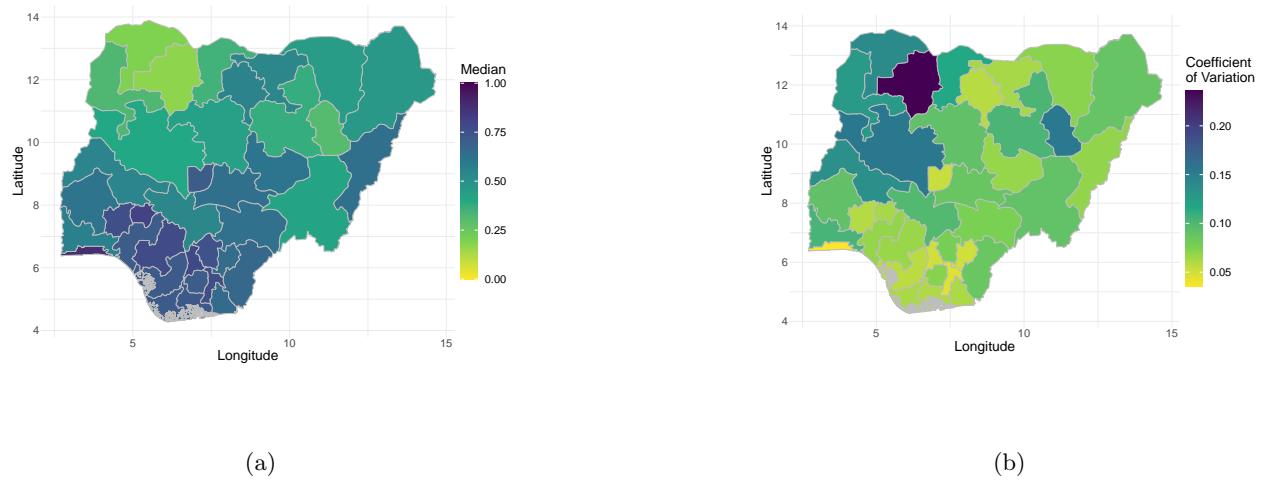


Figure 13: The median (a) and coefficient of variation (b) of 100 simulations of $P_a|\mathbf{Y}$, where \mathbf{X} is assumed to follow a Besag model and the precision parameter $\tau = \hat{\tau}$.