

TMA4250 Project 3: Gaussian Markov Random Fields

Ola Rasmussen & Maiken Hatten

April 24, 2024

Contents

1	Simulation and Visualization	2
2	Small Area Estimation	6

1 Simulation and Visualization

a)

For a Besag model, we have $x_i | \mathbf{x}_{-i}, \kappa \sim \mathcal{N} \left(\frac{1}{n_i} \sum_{j:j \sim i} x_j, \frac{1}{n_i \kappa} \right)$, where n_i denotes the number of neighbors of region i and $\kappa > 0$ is the precision parameter. This means that the precision matrix \mathbf{Q} can be written as $\mathbf{Q} = \kappa \mathbf{R}$, where \mathbf{R} is the structure matrix such that

$$R_{i,j} = \begin{cases} n_i & i = j, \\ -1 & i \sim j, \\ 0 & otherwise. \end{cases}$$

For our example, we then have $\mathbf{Q}_1 = \tau_1 \mathbf{R}_1$ and $\mathbf{Q}_2 = \tau_2 \mathbf{R}_2$, where $\tau_1 > 0$ and $\tau_2 > 0$ are the precision parameters, and \mathbf{R}_1 and \mathbf{R}_2 are the structure matrices. Using the neighborhood matrices, we first sum the rows, set the diagonal of the respective structure matrix as a vector containing these sums, and then subtracting the neighborhood matrices so the row sum is equal to zero. \mathbf{Q}_1 is a 37×37 matrix with rank $37 - 1 = 36$, and \mathbf{Q}_2 is a 775×775 matrix with rank $775 - 1 = 774$.

The sparsity pattern of \mathbf{Q}_1 and \mathbf{Q}_2 are plotted in Figure 1 and 2 respectively. The advantage of treating the Besag models as GMRFs instead of normal multivariate Gaussian distributions is that the sparsity of the precision matrices allows for faster computations.

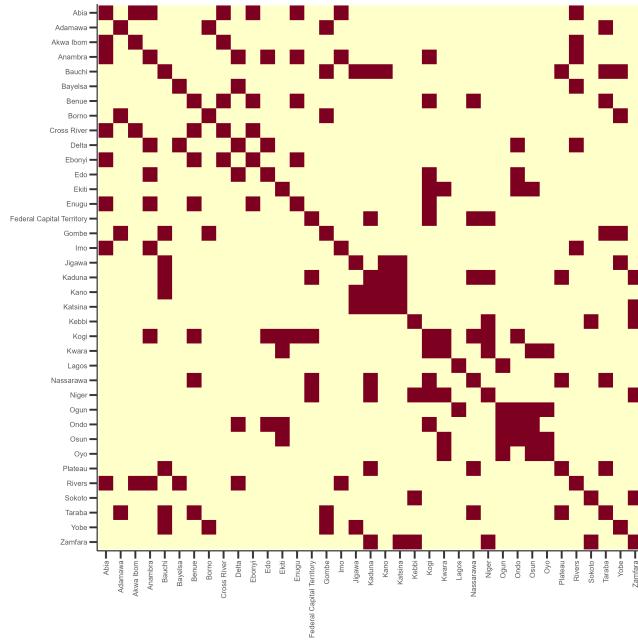


Figure 1: Sparsity pattern of \mathbf{Q}_1 .

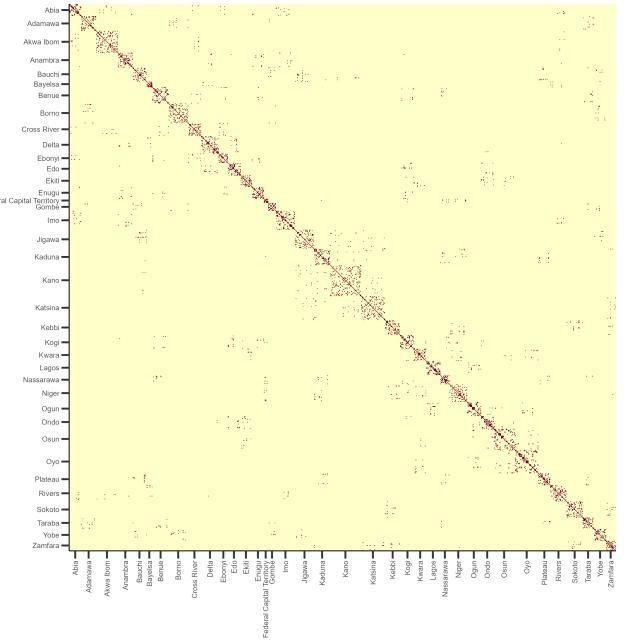


Figure 2: Sparsity pattern of \mathbf{Q}_2 .

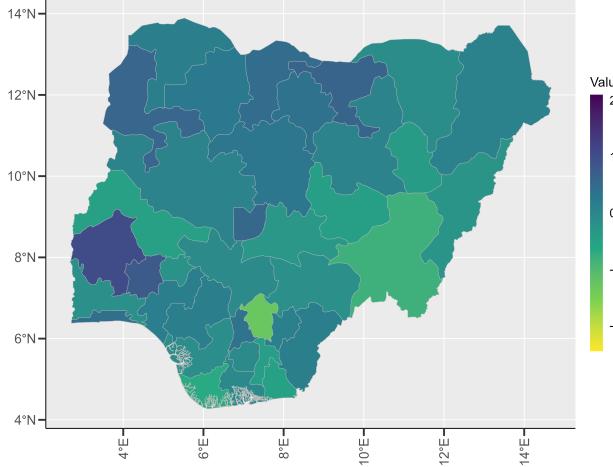
b)

The Besag model is an intrinsic GMRF of first order, so it is improper. However, an improper GMRF on its own cannot be sampled. Hence, if we want to obtain a sample, we need to sample with a sum-to-zero constraint in order to sample the proper component of the improper GMRF. The algorithms for sampling from a proper GMRF and the proper part of a first-order intrinsic GMRF are shown in Algorithm 1 and 2, respectively.

Algorithm 1 Sample from GMRF

- 1: **Set** μ mean
- 2: **Set** \mathbf{Q} Precision matrix
- 3: Compute Cholesky factor \mathbf{L} ($\mathbf{Q} = \mathbf{LL}^\top$)
- 4: Sample $\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$
- 5: Solve $\mathbf{L}^\top \mathbf{v} = \mathbf{z}$
- 6: Compute $\mathbf{x} = \mu + \mathbf{v}$
- 7:
- 8: **Return** $\mathbf{x} \sim \mathcal{N}(\mu, \mathbf{Q}^{-1})$

Figure 3 shows two realization from the Besag model on the admin1 graph with $\tau_1 = 1$ and sum-to-zero constraint, while Figure 4 shows two realization from $\mathcal{N}_{37}(\mathbf{0}, \mathbf{I}_{37})$. In the Besag model, due to the sum-to-zero constraint, the values generated in the realization are random around 0. The model's constraints also lead to a smaller range of values visible. Meanwhile, the multivariate normal model also has mean 0. In Figure 3 we observe that neighbouring areas tend to have similar values. This is expected, as the value of a given location is influenced by the values of its neighbours. However, Figure 4 does not show this same pattern. We observe that some neighbouring locations have distinctly different values, which suggests that the values for different areas are independent of each other.

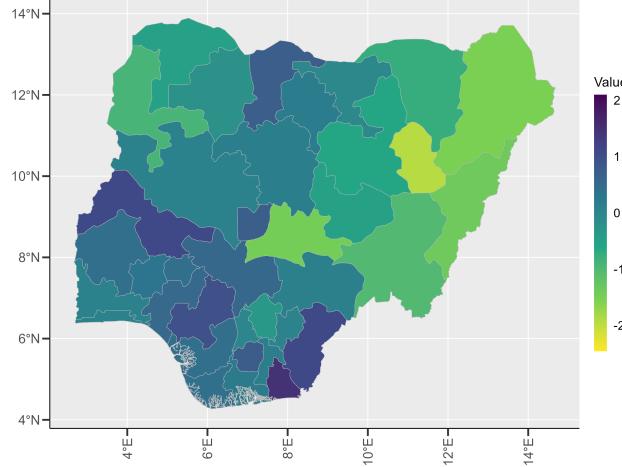


(a) First realization.

Algorithm 2 Sample from IGMRF

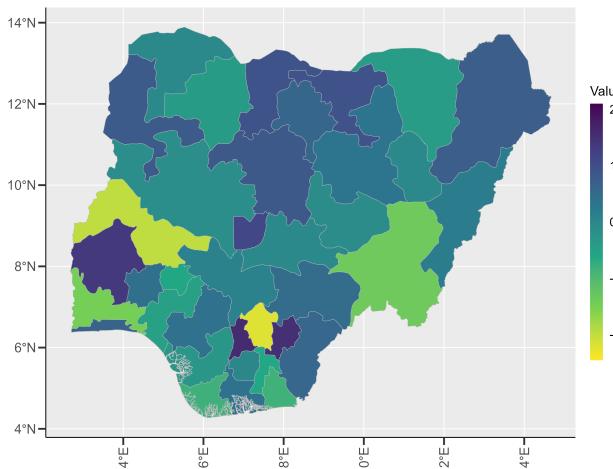
- 1: **Set** \mathbf{Q} Rank $n-1$ SPSD matrix s.t. all rows sum to zero
- 2: **Set** ϵ Small number
- 3: Set $\tilde{\mathbf{Q}} = \mathbf{Q} + \epsilon \mathbf{I}$
- 4: Compute Cholesky factor $\tilde{\mathbf{L}}$
- 5: Sample $\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$
- 6: Solve $\tilde{\mathbf{L}}^\top \mathbf{v} = \mathbf{z}$
- 7: Compute $\mathbf{x} = \mathbf{v} - \text{mean}(\mathbf{v}) \mathbf{1}$ ($\mathbf{1}$ is vector of ones)
- 8: **Return** $\mathbf{x} | \sum_{i=1}^n x_i = 0 \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}^{-1})$

on the admin1 graph with $\tau_1 = 1$ and sum-to-zero constraint, while Figure 4 shows two realization from $\mathcal{N}_{37}(\mathbf{0}, \mathbf{I}_{37})$.

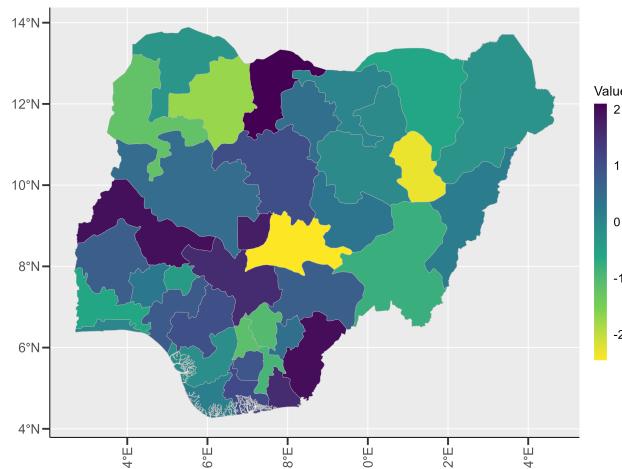


(b) Second realization.

Figure 3: Besag model on the admin1 graph with $\tau_1 = 1$, and sum-to-zero constraint.



(a) First realization.



(b) Second realization.

Figure 4: Multivariate Gaussian model with $\mu = \mathbf{0}$ and $\Sigma = \mathbf{I}$.

c)

Figure 5 shows two realizations with the Besag model on the admin2 graph with $\tau_2 = 1$, and sum-to-zero constraint, while Figure 6 shows one realization from $\mathcal{N}_{775}(\mathbf{0}, \mathbf{I}_{775})$. We see the same trends as in Section 1b), where the realization from the multivariate normal is shown to have greater distinctions between values between neighbours than the realization from the Besag model. Figure 5 shows that regions are still influenced by their neighbours, and this explains the lower variation in the values that the realization produces.

Meanwhile, the multivariate normal model does not account for this, so we again observe that each region's value is independent of its neighbours. These differences are more visible by the increased number of regions. This is because we observe more detailed patterns, which makes the Besag model's lower variance and the multivariate normal's randomness more prominent.

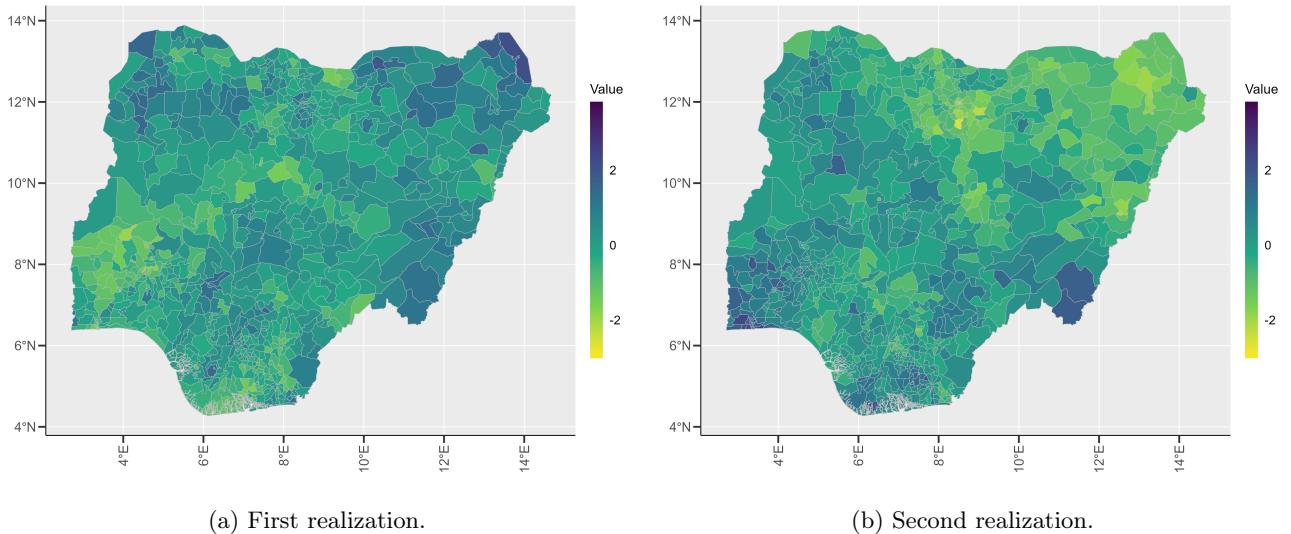


Figure 5: Besag model on the admin2 graph with $\tau_2 = 1$, and sum-to-zero constraint.

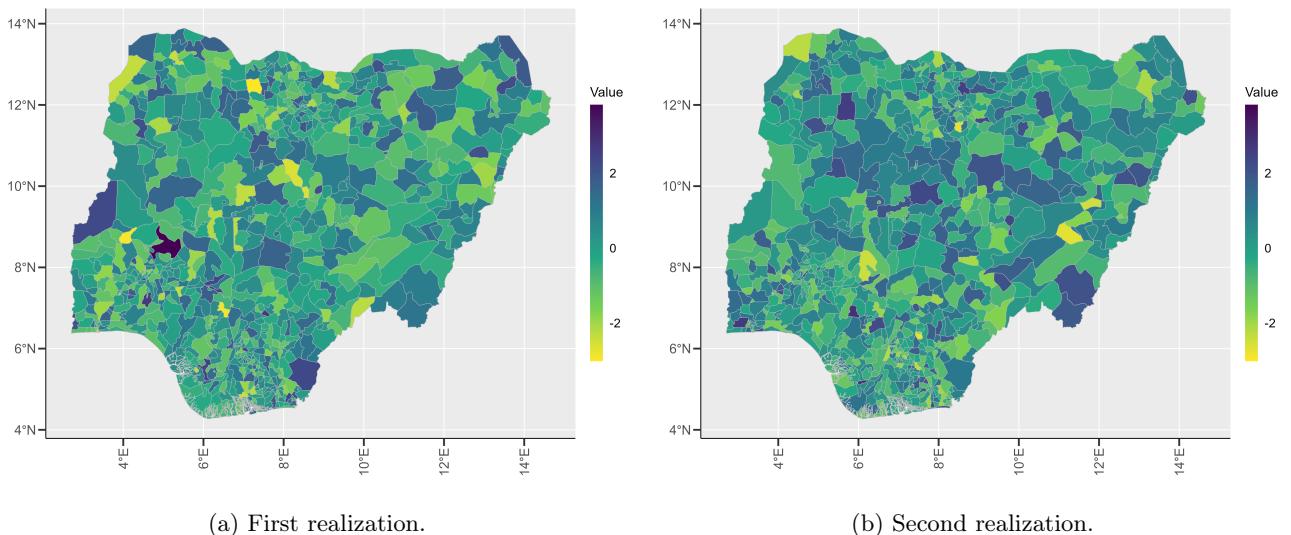


Figure 6: Multivariate Gaussian model with $\mu = \mathbf{0}$ and $\Sigma = \mathbf{I}$.

d)

Figure 7 shows the marginal variances in each admin2 area after generating 100 realizations from the Besag model. For a model to appear stationary, we need to observe both constant mean and variance. However, we observe in Figure 7 that the variance is not constant. Therefore the Besag model does not appear to be stationary.

Figure 8 shows the correlations between the admin2 area Gubio and all others after generating 100 realizations. Although the Besag model satisfies the pairwise Markov property, which states that for a GMRF,

$$x_i \perp x_j | \mathbf{x}_{-ij} \text{ if } \{i, j\} \neq \mathcal{E} \text{ and } i \neq j,$$

areas far apart can still be dependent because they aren't marginally independent.

Negative correlations arise as a result of the sum-to-zero constraint. As the constraint refers to $\sum x_i = 0$, this means that in order for positive correlations to exist, they must be offset by negative values.

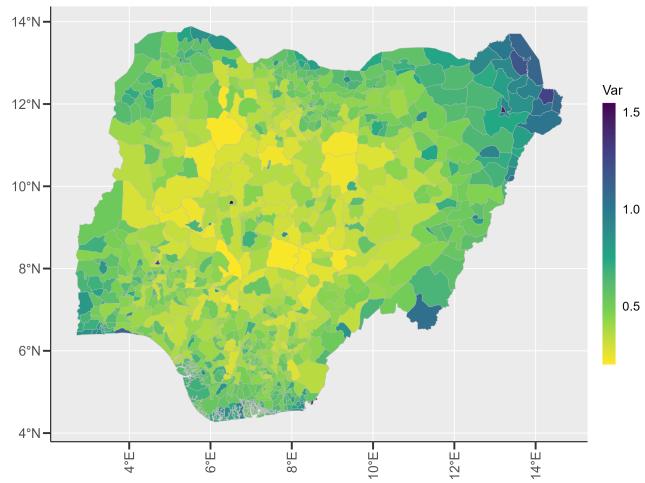


Figure 7: Empirical variance based on 100 realizations from the Besag model on the admin2 graph with $\tau_2 = 1$, and sum-to-zero constraint.

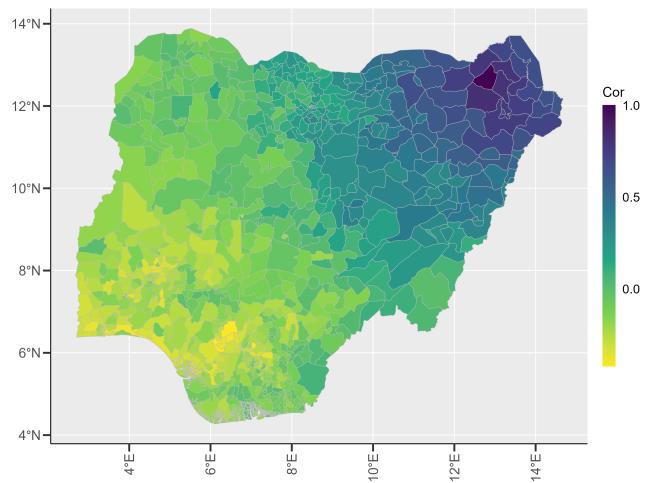


Figure 8: Empirical correlation between Gubio and the other locations based on 100 realizations from the Besag model on the admin2 graph with $\tau_2 = 1$, and sum-to-zero constraint.

2 Small Area Estimation

We consider the estimation of vaccination coverages for children in the 37 admin1 areas in Nigeria. We assume that there is no sharing of information between admin1 areas. Let p_a denote the true proportion of children who are vaccinated in area a for $a = 1, \dots, 37$. Let \hat{P}_a be the estimator for p_a that results from the above procedure. A common assumption is that

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

where $\text{logit}(x) = \log\left(\frac{x}{1-x}\right)$, V_1, \dots, V_{37} are known variances, and $\hat{P}_1, \dots, \hat{P}_{37}$ are independent. We will consider hierarchical spatial models where we imagine the true proportions to be stochastic variables. We use the notation $\mathbf{X} = (\text{logit}(P_1), \dots, \text{logit}(P_{37}))^\top$ and $\mathbf{Y} = (\text{logit}(\hat{P}_1), \dots, \text{logit}(\hat{P}_{37}))^\top$ for the random vector of true proportions and the random vector of observed proportions, respectively.

a)

The observed proportions, \hat{p}_a , are displayed in Figure 9. In terms of borrowing strength in space, we can refer to using information from surrounding areas of a region in order to reduce uncertainty in estimates. If Figure 9 showed a random pattern with no clear trends, this would imply that the spatial correlation is weak; this makes borrowing of strength less reasonable. However, Figure 9 shows clusters of high and low vaccination coverages, in the lower regions and upper regions, respectively. This implies that vaccination coverages in these clusters are influenced by similar factors. As a result, borrowing strength in space can lead to reducing uncertainty, so it seems reasonable.

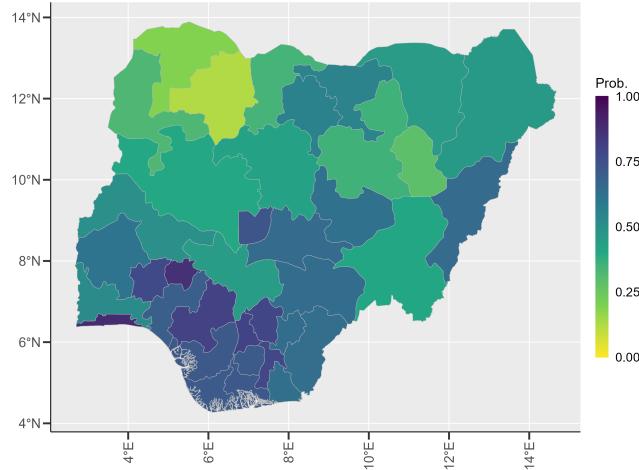


Figure 9: Observed proportions $\hat{\mathbf{p}}$.

b)

We have the model $\mathbf{Y} = \mathbf{X} + \boldsymbol{\epsilon}$, where $\boldsymbol{\epsilon} = (\epsilon_1, \dots, \epsilon_{37})^\top \sim \mathcal{N}_{37}(\mathbf{0}, \mathbf{V})$ and

$$\mathbf{V} = \begin{bmatrix} V_1 & 0 & \cdots & 0 \\ 0 & V_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & V_{37} \end{bmatrix}.$$

The distribution of $\mathbf{Y} | \mathbf{X}$ is then multivariate Gaussian with mean vector $\boldsymbol{\mu} = \mathbf{X}$ and covariance matrix \mathbf{V} .

We are given the prior $\mathbf{X} \sim \mathcal{N}_{37}(\mathbf{0}, \sigma^2 \mathbf{I}_{37})$, where $\sigma^2 = 100^2$. In the lectures we have shown that if we have the hierarchical spatial model

$$\begin{aligned}\mathbf{Y} | \mathbf{X} &\sim \mathcal{N}_n(\mathbf{MX}, \mathbf{Q}_2^{-1}), \quad \mathbf{M} \in \mathbb{R}^{n \times m}, \\ \mathbf{X} &\sim \mathcal{N}_m(\mathbf{0}, \mathbf{Q}_1^{-1}), \\ \Rightarrow \mathbf{X} | \mathbf{Y} &\sim \mathcal{N}_m([\mathbf{Q}_1 + \mathbf{M}^\top \mathbf{Q}_2 \mathbf{M}]^{-1} \mathbf{M}^\top \mathbf{Q}_2 \mathbf{Y}, [\mathbf{Q}_1 + \mathbf{M}^\top \mathbf{Q}_2 \mathbf{M}]^{-1}).\end{aligned}$$

In our case, $n = 37$, $m = 37$, $\mathbf{M} = \mathbf{I}$, $\mathbf{Q}_1 = \frac{1}{\sigma^2} \mathbf{I}$, and $\mathbf{Q}_2 = \mathbf{V}^{-1}$, so

$$\mathbf{X} | \mathbf{Y} \sim \mathcal{N}_{37}\left(\left[\frac{1}{\sigma^2} \mathbf{I} + \mathbf{V}^{-1}\right]^{-1} \mathbf{V}^{-1} \mathbf{Y}, \left[\frac{1}{\sigma^2} \mathbf{I} + \mathbf{V}^{-1}\right]^{-1}\right).$$

When $\sigma^2 \rightarrow \infty$ the expected value will converge to \mathbf{Y} and the covariance matrix will converge to \mathbf{V} , and since $X_a = \text{logit}(P_a)$, so $P_a | \mathbf{Y} = \mathbf{y} \sim \text{expit}(\mathcal{N}_{37}(\mathbf{y}, \mathbf{V}))$.

The empirical median and the empirical coefficient of variation for $P_a | \mathbf{Y} = \mathbf{y}$ based on 100 samples are displayed in Figure 10 and Figure 11, respectively. The estimated medians are similar to the estimated values displayed in Figure 9. The coefficient of variation is low across the whole map. Lower values imply more precise estimates.

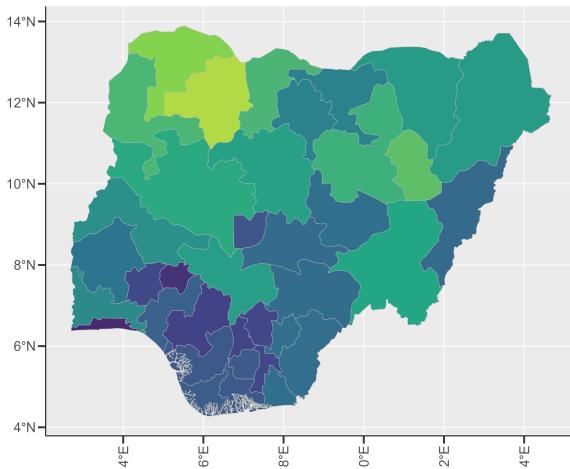


Figure 10: Empirical median for $P_a | \mathbf{Y} = \mathbf{y}$ based on 100 samples.

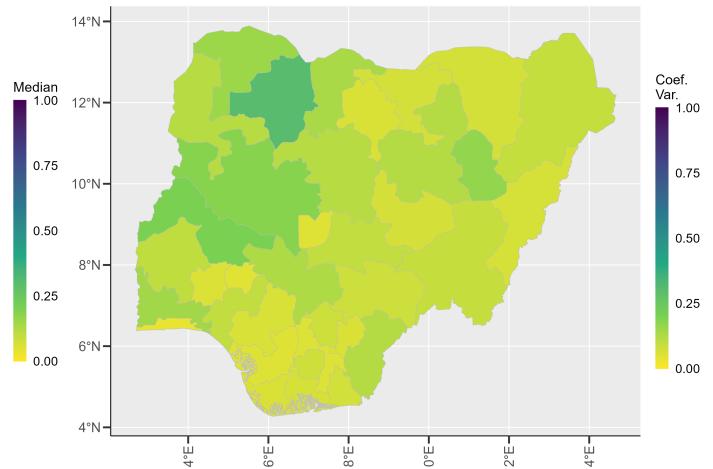


Figure 11: Empirical coefficient of variation for $P_a | \mathbf{Y} = \mathbf{y}$ based on 100 samples.

c)

Now assuming \mathbf{X} a priori follows a Besag model, defined on the admin1 area graph, with precision parameter τ . Using the result from the lectures, and $n = 37$, $m = 37$, $\mathbf{Q}_1 = \tau \mathbf{R}_1$, $\mathbf{M} = \mathbf{I}$, and $\mathbf{Q}_2 = \mathbf{V}^{-1}$, we get

$$\mathbf{X} | \mathbf{Y} \sim \mathcal{N}_{37}\left([\tau \mathbf{R}_1 + \mathbf{V}^{-1}]^{-1} \mathbf{V}^{-1} \mathbf{Y}, [\tau \mathbf{R}_1 + \mathbf{V}^{-1}]^{-1}\right).$$

The empirical median and the empirical coefficient of variation for the new model $P_a | \mathbf{Y} = \mathbf{y}$ based on 100 samples with $\tau = 1$ are displayed in Figure 12 and Figure 13, respectively. The estimated median have less variation between close areas when compared to Figure 9 and Figure 10. The coefficient of variation is lower than the coefficient of variation in Figure 11.

To determine whether the GMRF is intrinsic or proper, we note that a proper GMRF has full-rank. In our case, the precision matrix has rank 37. An intrinsic GMRF is an improper GMRF of rank $n - 1$, so our GMRF is proper and not intrinsic.

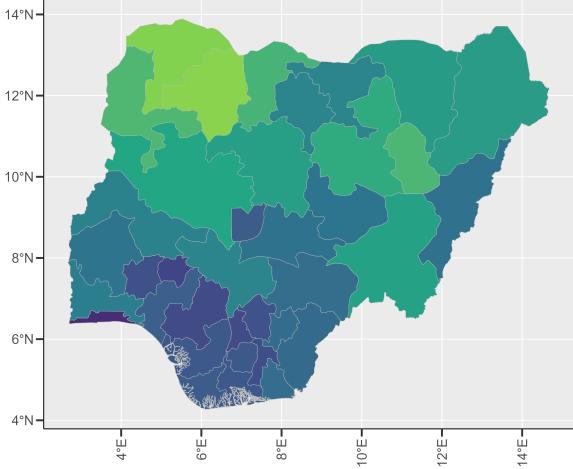


Figure 12: Empirical median for $P_a | \mathbf{Y} = \mathbf{y}$ based on 100 samples with $\tau = 1$.

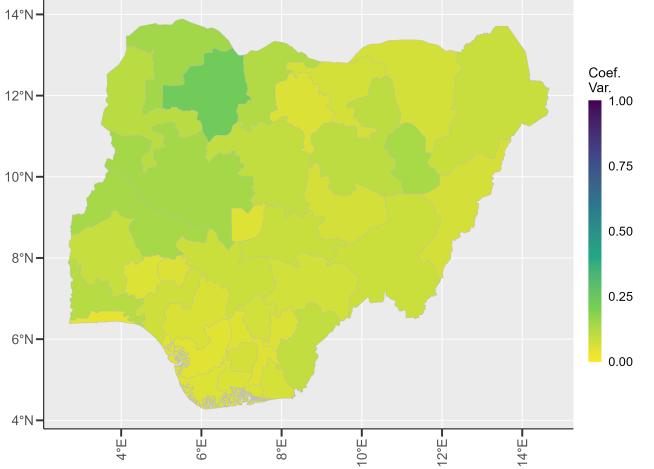


Figure 13: Empirical coefficient of variation for $P_a | \mathbf{Y} = \mathbf{y}$ based on 100 samples with $\tau = 1$.

d)

Now we let $Y_{38} | P_{\text{Kaduna}} \sim \mathcal{N}(\text{logit}(P_{\text{Kaduna}}), 0.1^2)$, where $Y_{38} | \mathbf{P}$ is independent of $\mathbf{Y} | \mathbf{P}$. Let $\tilde{\mathbf{Y}} = (Y_1, \dots, Y_{37}, Y_{38})^\top$. Now $n = 38$, $m = 37$, $\mathbf{Q}_1 = \tau \mathbf{R}_1$,

$$\mathbf{M} = \begin{bmatrix} 1 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \cdots & \cdots & \cdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 & 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 1 \\ 0 & 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \end{bmatrix}$$

where the 1 in the last row is the Kaduna position, and

$$\mathbf{Q}_2 = \begin{bmatrix} \mathbf{V} & 0 \\ 0 & 0.1^2 \end{bmatrix}.$$

We then have

$$\mathbf{X} | \tilde{\mathbf{Y}} \sim \mathcal{N}_{37} \left([\tau \mathbf{R}_1 + \mathbf{M}^\top \mathbf{Q}_2 \mathbf{M}]^{-1} \mathbf{M}^\top \mathbf{Q}_2 \tilde{\mathbf{Y}}, [\tau \mathbf{R}_1 + \mathbf{M}^\top \mathbf{Q}_2 \mathbf{M}]^{-1} \right).$$

The empirical median and the empirical coefficient of variation for the new model $P_a | \tilde{\mathbf{Y}} = \tilde{\mathbf{y}}$ based on 100 samples with $y_{38} = 0.5$ and $\tau = 1$ are displayed in Figure 14 and Figure 15, respectively. Figure 15 and Figure 13 show similar empirical coefficients of variation which are lower than what is observed in Figure 11. This trend extends to the surrounding regions of Kaduna in Problems b) and c). This shows how surrounding regions are influenced by the values for Kaduna. Meanwhile, the empirical median found in Figures 10, 12, and 14 remain very similar, apart from the change in empirical median in some sections of Figure 12 and Figure 14 (such as the upper left corner of the map). This may be due to the introduction of borrowing strength from neighboring areas. Conversely, the other regions with greater similarity may be attributed to how vaccination coverage patterns may remain the same in different geographical regions, regardless of different approaches to modeling.

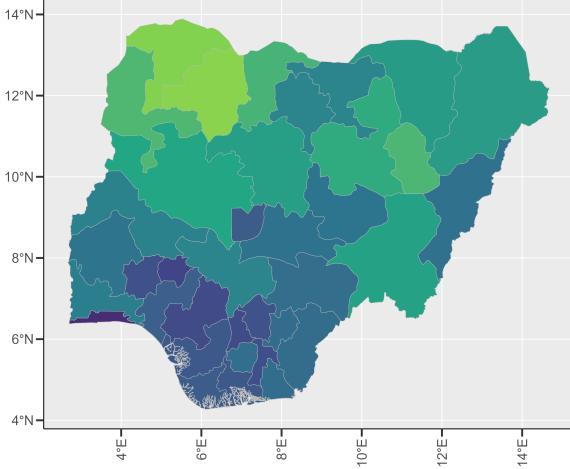


Figure 14: Empirical median for $P_a | \tilde{\mathbf{Y}} = \tilde{\mathbf{y}}$ based on 100 samples with $y_{38} = 0.5$ and $\tau = 1$.

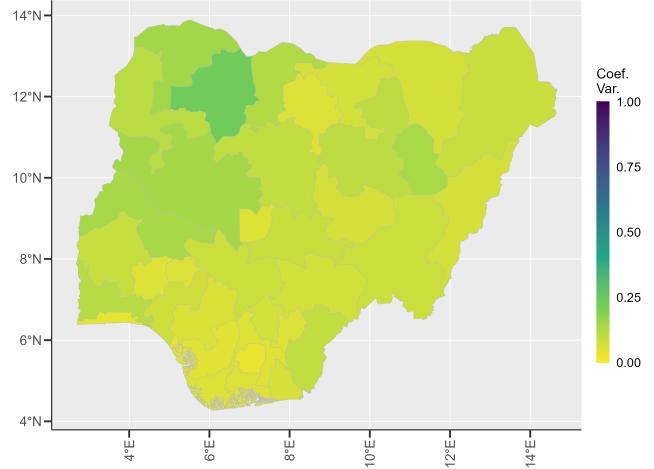
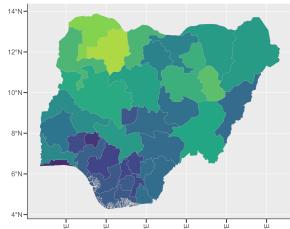


Figure 15: Empirical coefficient of variation for $P_a | \tilde{\mathbf{Y}} = \tilde{\mathbf{y}}$ based on 100 samples with $y_{38} = 0.5$ and $\tau = 1$.

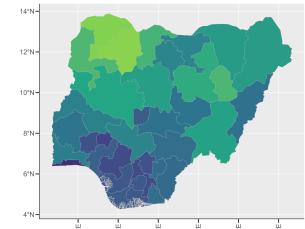
e)

In this part, we will investigate the sensitivity of τ . We have repeated what we did in Section 2c), but now with $\tau = 0.1$ and $\tau = 10$. The empirical median and the empirical coefficient of variation for the new model $P_a | \mathbf{Y} = \mathbf{y}$ based on 100 samples with three different values of τ are displayed in Figure 16 and Figure 17, respectively. We can see that when τ increases, the variation between areas decreases. This is a result of estimates being more influenced by their neighbors as τ increases.

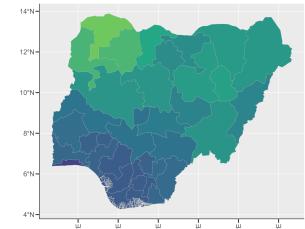
Increasing τ also leads adjacent areas having greater influence on each other, with Figure 17c showing this effect more visibly than when $\tau = 0.1$ or $\tau = 1$. This shows that the different values of τ lead to different results, and the larger the differences between τ , the more apparent these differences are. This suggests that the medians and coefficients of variation are considerably sensitive with respect to τ .



(a) $\tau = 0.1$.

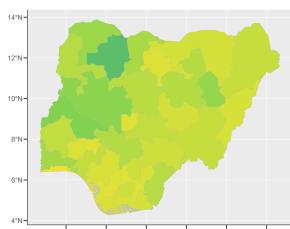


(b) $\tau = 1$.

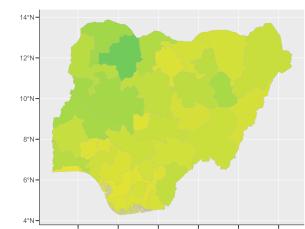


(c) $\tau = 10$.

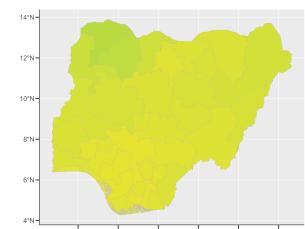
Figure 16: Empirical median for $P_a | \mathbf{Y} = \mathbf{y}$ based on 100 samples with three different values of τ .



(a) $\tau = 0.1$.



(b) $\tau = 1$.



(c) $\tau = 10$.

Figure 17: Empirical coefficient of variation for $P_a | \mathbf{Y} = \mathbf{y}$ based on 100 samples with three different values of τ .

f)

To find the log-likelihood for the model in 2, we note that this log-likelihood holds for any $\mathbf{x} \in \mathbb{R}^{37}$. We apply Bayes' theorem to find that

$$\begin{aligned} l(\tau; \mathbf{y}) &= \log(f(\mathbf{y}; \tau)) = \log\left(\frac{f(\mathbf{x}; \tau)f(\mathbf{y}|\mathbf{x})}{f(\mathbf{x}|\mathbf{y}; \tau)}\right) \\ &= \log(f(\mathbf{x}; \tau)) + \log(f(\mathbf{y}|\mathbf{x})) - \log(f(\mathbf{x}|\mathbf{y}; \tau)). \end{aligned}$$

We know from the lecture notes that in general, for a GMRF \mathbf{X} with mean μ and precision matrix $\mathbf{Q} > 0$,

$$\log f(\mathbf{x}) = -\frac{n}{2} \log(2\pi) + \sum_{i=1}^n \log L_{ii} - \frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^\top \mathbf{Q}(\mathbf{x} - \boldsymbol{\mu}).$$

Given that $\sum_{i=1}^n L_{ii} = (1/2) \log |\mathbf{Q}|$, we also know from earlier that $\mathbf{Y}|\mathbf{X}$ has mean vector $\boldsymbol{\mu} = \mathbf{X}$ and covariance matrix \mathbf{V} , $\mathbf{X}|\mathbf{Y}$ has mean $[\tau \mathbf{R}_1 + \mathbf{V}^{-1}] \mathbf{V}^{-1} \mathbf{Y}$ and covariance matrix $[\tau \mathbf{R}_1 + \mathbf{V}^{-1}]^{-1}$, and $\mathbf{X} \sim \mathcal{N}_m(\mathbf{0}, \mathbf{Q}_1^{-1})$, where we define \mathbf{Q}_1 as $\tau \mathbf{R}_1$. Using τ as our precision parameter and substituting notation such that the covariance matrix \mathbf{V} is now \mathbf{D} , we find that

$$\begin{aligned} l(\tau; \mathbf{y}) &= \text{Const} + \frac{37-1}{2} \log(\tau) - \frac{1}{2} \log |\mathbf{Q}| + \frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^\top \mathbf{Q}(\mathbf{x} - \boldsymbol{\mu}) \\ &\quad - \frac{1}{2}(\mathbf{x} - \mathbf{0})^\top \tau \mathbf{R}(\mathbf{x} - \mathbf{0}) - \frac{1}{2}(\mathbf{y} - \mathbf{x})^\top \mathbf{D}^{-1}(\mathbf{y} - \mathbf{x}) \\ &= \text{Const} + \frac{37-1}{2} \log(\tau) - \frac{1}{2} \log |\mathbf{Q}| + \frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^\top \mathbf{Q}(\mathbf{x} - \boldsymbol{\mu}) \\ &\quad - \frac{\tau}{2} \mathbf{x}^\top \mathbf{R} \mathbf{x} - \frac{1}{2}(\mathbf{y} - \mathbf{x})^\top \mathbf{D}^{-1}(\mathbf{y} - \mathbf{x}). \end{aligned}$$

For the MLE $\hat{\tau}$ of τ , we estimate that this value is 0.806. As a result, Figure 18 and 19 show very similar results as the results found in c). This suggests that we can use an estimate between 0.1 and 1 that leads to a lower variance and greater borrowing of strength than if we were to use $\tau = 1$, and the opposite case if we were to use $\tau = 0.1$. This is indicative of the bias-variance tradeoff.

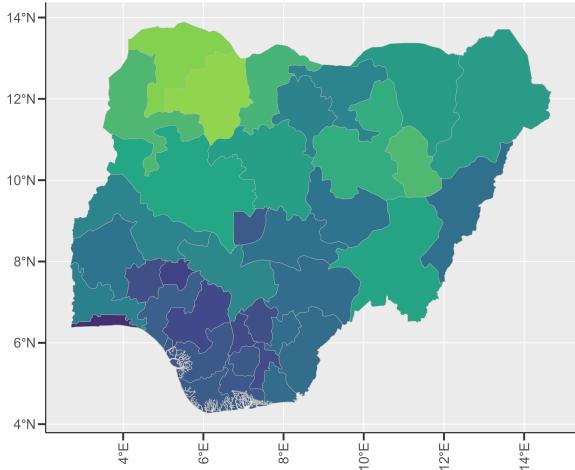


Figure 18: Empirical median for $P_a | \mathbf{Y} = \mathbf{y}$ based on 100 samples with $\tau = \hat{\tau}$.

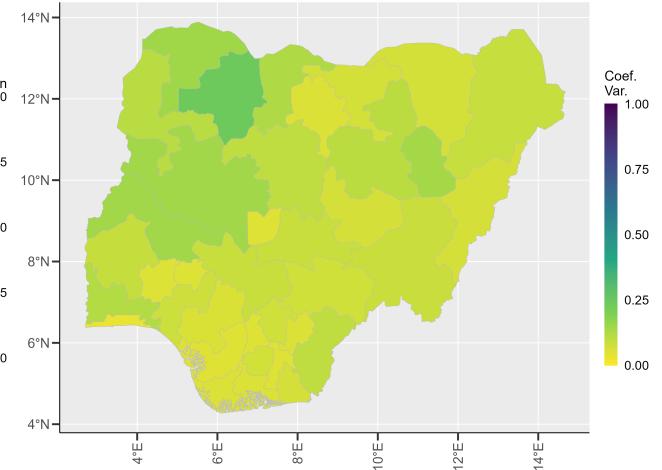


Figure 19: Empirical coefficient of variation for $P_a | \mathbf{Y} = \mathbf{y}$ based on 100 samples with $\tau = \hat{\tau}$.