TMA4250 - Spatial Statistics

Project 3: Gaussian Markov Random Fields

Authors:
Gard Gravdal & Marcus Hjørund

Date: 04.05.2024

1 Introduction

In this project, we will be working with problems related to Gassian Markov random fields (GMRFs). We will be working with the geography of Nigeria, which has two nested subdivisions. The first administrative level is called *admin1*, which consists of 37 areas. The second administrative level is called *admin2*, which consists of 775 areas. We will be working with several files in this project, namely:

- *Admin1Geography.RData*: Contains an object nigeriaAdm1 that contains the borders of the 47 admin1 areas
- *Admin2Geography.RData*: Contains an object nigeriaAdm2 that contains the borders of the 775 admin2 areas
- *Admin1Graph.txt*: a 37×37 matrix giving the admin1 neighbourhood structure
- *Admin2Graph.txt*: a 775×775 matrix giving the admin2 neighbourhood structure
- *DirectEstimates.txt*: Used in Problem 2.

2 Problem 1: Simulation and Visualization

In this problem we will use the term admin1 graph to refer to the graph structure arising from connecting admin1 areas that share a border, and the term admin2 graph to refer to the graph structure arising from connecting admin2 areas that share a border. These graphs are found in Admin1Graph.txt and Admin2Graph.txt.

2.1 a)

In this section we will describe how to use the neighbourhood matrices to construct the precision matrices of the *Besag model* on the admin graphs. We denote the admin1 graph by $\mathbf{Q}_1 = \tau_1 \mathbf{R}_1$ and the admin2 graph by $\mathbf{Q}_2 = \tau_2 \mathbf{R}_2$. Here $\tau_1, \tau_2 > 0$ denote the precision parameters and $\mathbf{R}_1, \mathbf{R}_2$ denote the structure matrices.

We begin by introducing some terminology. For neighbourhood relations, we have:

- $\nu = \{1, 2, ..., n\}$: Finite set of nodes
- $\sim$: Neighbourhood relation, $i \sim j$ implies i and j are neighbours
- $\partial i \subseteq \nu$: Neighbours of node i
- $(\nu, \sim)$: Neighbourhood structure

For a graph $G = (\nu, \epsilon)$, we define the set of neighbours of i as $\text{ne}(i) = \{j \in \nu : \{i, j\} \in \epsilon\}$ for $i \in \nu$, where ν are nodes and ϵ denotes the edges, such that $i \sim j \Leftrightarrow \{i, j\} \in \epsilon$.

Next, we can define the Besag model as

Definition 2.1 (Besag model)**.** The improper GMRF $\boldsymbol{X} = (X_1, ..., X_n)^T$ with respect to the connected graph $G = (\nu, \epsilon)$ is called a Besag model with precision parameter $\kappa > 0$ if

$$f(\boldsymbol{x}; \kappa) \propto \kappa^{\frac{n-1}{2}} \exp\left(-\frac{\kappa}{2} \sum_{i \sim j} (x_i - x_j)^2\right), \quad \boldsymbol{x} \in \mathbb{R}^n \tag{1}$$

The Besag model is used to model the spatial similarity. Note that a first-order random walk is a Besag model on a linear graph. In other words, we have that $x_i - x_j \sim \mathcal{N}(0, \sigma^2)$ if $i \sim j$. Intuitively, the Besag model says that things nearby are similar, and that the value at neighbouring sites is probably not more than 3σ apart (BATH 2024). Next we want to use the neighbourhood matrices to construct the precision matrices of the Besag model on the admin graphs. To construct the precision matrices, we use the fact that

$$
\begin{aligned}
f(\boldsymbol{x}; \kappa) &\propto \kappa^{\frac{n-1}{2}} \exp\left(-\frac{\kappa}{2}\sum_{i\sim j}(x_i - x_j)^2\right) && (2)\\
&= \kappa^{\frac{n-1}{2}} \exp\left(-\frac{\kappa}{2}\sum_{i\sim j} x_i^2 - 2x_i x_j + x_j^2\right) && (3)
\end{aligned}
$$

The precision matrices can be found by matching with the quadratic term from the general GMRF:

$$
\begin{aligned}
&\exp\left(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T \mathbf{Q}(\boldsymbol{x}-\boldsymbol{\mu})\right)\\
&= \exp\left(-\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} x_i Q_{ij} x_j\right)
\end{aligned}
$$

Where we have used that $\boldsymbol{\mu} = 0$, since there are no constant terms in equation (3). Further, we see that for $i = j$, we have $x_i Q_{ii} x_i = \kappa \sum_{i\sim j} x_i^2 \Rightarrow Q_{ii} = \kappa|\mathrm{ne}(i)|$. When $i \sim j$, we find that $Q_{ij} = -\kappa$, since we match the terms $x_i Q_{ij} x_j + x_j Q_{ji} x_i = -2\kappa x_i x_j$, and we use the fact that $\mathbf{Q}$ is symmetric. The remaining elements of $\mathbf{Q}$ are 0.

We use the result just derived to find the precision matrices of the Besag model on the admin1 graph and admin2 graph. To write the matrices, we make use of the symmetric helper function $s(i, j)$ defined as

$$
s(i,j) = \begin{cases} -1, & i \sim j \\ 0, & \text{otherwise} \end{cases}
$$

Then, the precision matrices are given by

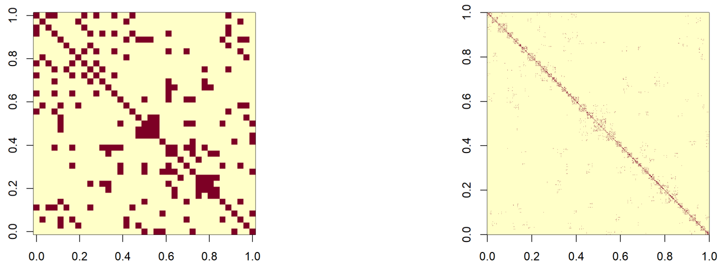
$$
\mathbf{Q}_k = \kappa \begin{pmatrix} |\mathrm{ne}(1)| & s(1,2) & s(1,3) & \dots & s(1,n_k) \\ s(2,1) & |\mathrm{ne}(2)| & s(2,3) & \dots & \\ \vdots & & \ddots & & \vdots \\ s(n_k,1) & & \dots & & |\mathrm{ne}(n_k)| \end{pmatrix} = \kappa \mathbf{R}_k \tag{4}
$$

where $k \in \{1, 2\}$, $\tau_1 = \tau_2 = \kappa$ and $n_1, n_2 = 37, 775$. The precision matrices are of dimension $n_k \times n_k$ and of rank $n_k - 1$. We find that it is of rank $n_k - 1$ since we can determine the i'th row if we know the other $n_k - 1$ rows, showing linear dependence between the rows. In the i'th row, the i'th element (the diagonal element of the matrix), is equal to the sum of the other elements of the i'th column. The other elements of the row can be found by the symmetry of the matrix.

Next we compute the proportion of non-zero elements numerically, and then we display the sparsity pattern for each of the precision matrices. For the admin1 precision matrix we have $\approx 15.27\%$ non-zero elements. In contrast, the admin2 precision matrix has $\approx 0.88\%$ non-zero elements. The sparsity patterns for each matrix can be seen in Figure 1, where we can clearly see the difference in sparsity between the matrices. The main benefit of treating the Besag models as GMRFs is that we get the sparsity we can see in Figure 1. When we are dealing with sparse matrices we can greatly speed up computations. If we are treating the Besag models as standard multivariate Gaussian distributions we obtain dense matrices, so by using GMRFs we are reducing complexity in operations from $\mathcal{O}(n^3)$ to around $\mathcal{O}(n^{1.5})$.

2.2 b)

In this section we will simulate from GMRF-models, and then generate two realizatons from the distributions. The first model we simulate from is the Besag model. When simulating the Besag



(a) admin1 (b) admin2

Figure 1: Sparsity patterns for the admin1 and admin2 administrative levels. Red indicates non-zero entry, yellow indicates zero entry.

model, we use the fact the Besag model is an intrinsic improper GMRF, with precision matrix $\mathbf{Q}$ of rank $n-1$.

Definition 2.2 (Intrinsic GMRF)**.** An *intrinsic* GMRF of first-order is an improper GMRF of rank $n-1$ where

$$\mathbf{Q}\begin{bmatrix}1\\ \vdots\\ 1\end{bmatrix} = 0.$$

In order to simulate from a proper distribution we use that fact that we can partition our data $\boldsymbol{X}$ into a "proper" and "improper" part

$$\boldsymbol{X} = \boldsymbol{X}^{\parallel} + \boldsymbol{X}^{\perp}$$

where $\boldsymbol{X}^{\parallel} \in \text{Null}(\mathbf{Q})$ and $\boldsymbol{X}^{\perp} \in \text{Null}(\mathbf{Q})^{\perp}$, $\text{Null}(\cdot)$ denoting the null-space of $\mathbf{Q}$. Since the Besag model is intrinsic,

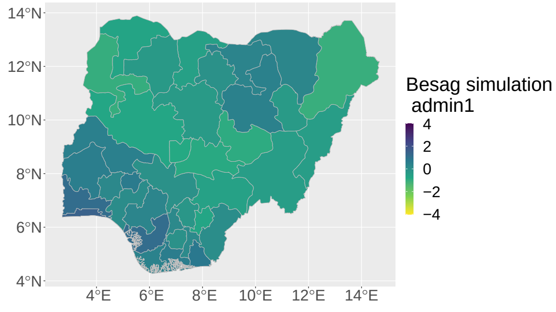
$$\text{Null}(\mathbf{Q}) = \left\{ \begin{bmatrix}1\\ \vdots\\ 1\end{bmatrix} c : c \in \mathbb{R} \right\}$$

and we can see that $\boldsymbol{X}$ has a flat distribution on level $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$, since the mean is a scalar number. Thus we can model $\boldsymbol{X}^{\perp} = \boldsymbol{X} - \mathbf{1}_n\bar{X}$ which is mapping the improper part of $\boldsymbol{X}$ to the "proper" part $\boldsymbol{X}^{\perp}$, and where $\mathbf{1}_n$ represents the vector of ones, size n. Subtracting the mean from each realization is equivalent to having a sum-to-zero constraint when simulating, and this ensures that we are sampling from the proper part of $\boldsymbol{X}$. The algorithm for simulation is

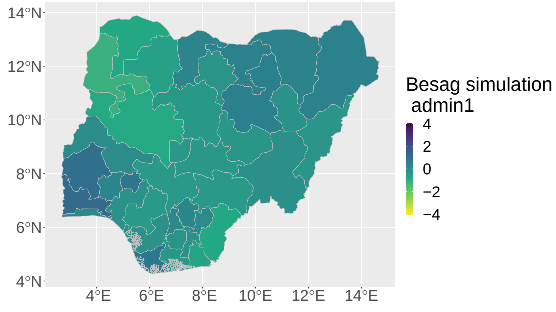
Set $\tilde{\mathbf{Q}} \leftarrow \mathbf{Q} + \epsilon\mathbf{I}_n$
Compute Cholesky factor $\tilde{\boldsymbol{L}}$
Sample $\boldsymbol{z} \sim \mathcal{N}_n(\mathbf{0}, \mathbf{I}_n)$
Solve $\tilde{\boldsymbol{L}}^T \boldsymbol{v} = \boldsymbol{z}$
Compute $\boldsymbol{x} = \boldsymbol{v} - \text{mean}(\boldsymbol{v})\begin{bmatrix}1\\ \vdots\\ 1\end{bmatrix}$

where $\epsilon \in \mathbb{R}^{+}$ is small when compared to the elements of $\mathbf{Q}$. $\mathbf{Q}$ is constructed using the method shown in Section 2.1. We add ϵ so that the matrix $\tilde{\mathbf{Q}}$ has full rank, which leaves a small error, however the method works well for small ϵ.

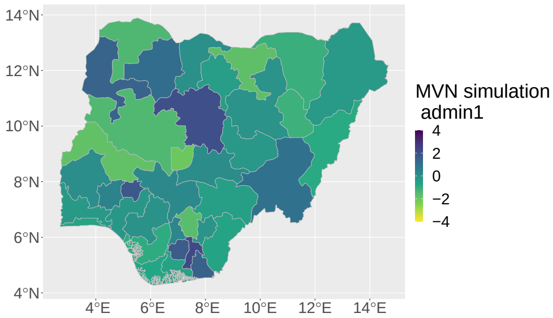
We want to compare the Besag model to the multivariate standard normal distribution, $\boldsymbol{z} \sim \mathcal{N}_n(\mathbf{0}, \mathbf{I}_n)$ to see which is better suited. We first compare the models using the admin1 graph, with precision parameter $\tau = 1$. From Figure 2 we can see simulations of the admin1 administrative division using the Besag model and a standard multivariate normal model. As expected we can



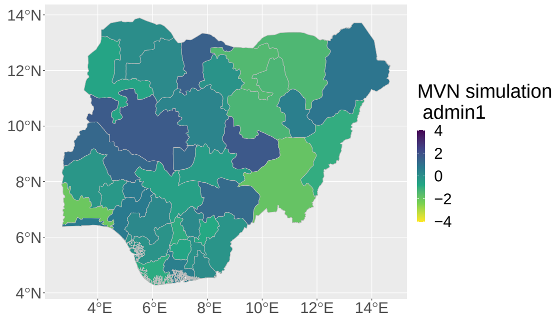
(a) Besag simulation $\tau = 1$



(b) Besag simulation $\tau = 1$



(c) Standard MVN simulation



(d) Standard MVN simulation

Figure 2: Simulation using the Besag and standard MVN models for admin1 administrative division

see that the areas in close proximity are affected by each other in the Besag model, while the areas from the multivariate normal model do not show any spatial structure.

2.3 c)

We now want to examine the same using the admin2 administrative division.

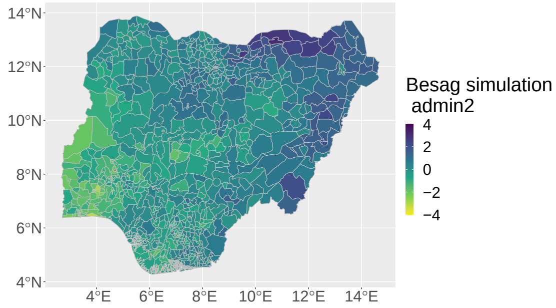
From Figure 3 we can see that in the case of much smaller administrative areas that we see the same trends as we saw in the case of the admin1 division. We have in the Besag model spatial correlation between the areas, while in the standard multivariate normal case we see that areas obtain values independently of its neighbours. It is not clear whether the differences between the distributions are more clear for the admin2 graph than the admin1 graph.

2.4 d)

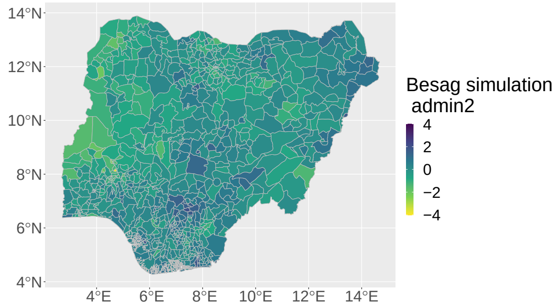
Next, we want to generate 100 realizations from the Besag model on the admin2 graph with $\tau_2 = 1$ and sum-to-zero constraint. We want to answer the following question: does the Besag model appear to be stationary?

Looking at Figure 4, we see that the areas in the centre of Nigeria have less variance than the areas close the the borders. This is because the presence of neighbours reduces variance as we have a model with spatial covariance. Thus, the model appears to not be stationary, since it has spatially varying marginal variance.

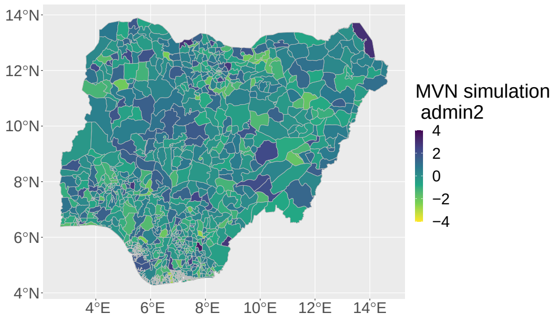
We will now consider one area in the admin2 division, namely the area *Gubio*. Even though the Besag model satisfies the pairwise Markov property, we still have correlation even though the areas are far apart, which can be seen from Figure 5. This is because the Markov property states that given all the areas, the value in Gubio will only be directly influenced by its neighbours. This is because of the flow of information, if we have the neighbours of Gubio, no further information is gained by conditioning on the neighbours further apart. But if we are not given the neighbours in



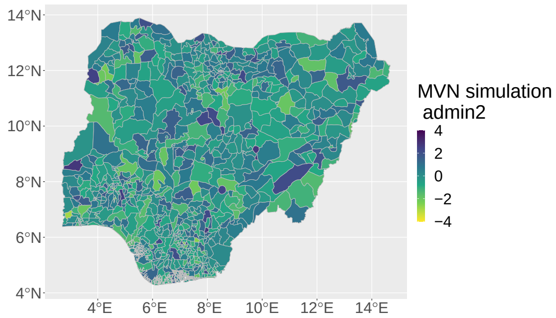
(a) Besag simulation $\tau = 1$



(b) Besag simulation $\tau = 1$



(c) Standard MVN simulation



(d) Standard MVN simulation

Figure 3: Simulation using the Besag and standard MVR models for admin2 administrative division

closest proximity, we will have dependence between Gubio and the areas further away. The Markov property does not state that the areas are independent, but rather that given all areas, only the neighbours give information. We see that some states have negative correlation with Gubio. When states that are neighbours are similar, states beyond a certain number of neighbours will tend to be dissimilar, leading to negative correlation between them. We see that the more neighbours there are between a state and Gubio, the more likely we are to have a negative correlation. This pattern would be more clear if we ran with more than 100 simulations.

3 Problem 2: Small Area Estimation

In this problem we will consider the estimation of vaccination coverages (proportion vaccinated) for children in the 37 admin1 areas in Nigeria. Nigeria does not have a complete registartion system for vaccines, which makes it challenging. A huge amount of the effort is used to coordinate the collection of a sample of children in each admin1 ares in such a way that this sample can be used to produce estimates of vaccination coverages for the admin1 areas. We will not consider details for how this is done in this project.

We let p_a denote the true proportion of children who are vaccinated in area a for $a = 1, ..., 37$. We let $\hat{P}_a$ be the estimator for p_a that results from the procedure described previously. We make the assumption that

$$\operatorname{logit}(\hat{P}_a) \sim \mathcal{N}(\operatorname{logit}(p_a), V_a), \quad a = 1, ..., 37 \tag{5}$$

where $V_1, ..., V_{37}$ are known variances, and $\hat{P}_1, ..., \hat{P}_{37}$ are independent. We want to learn about $\boldsymbol{p} = (p_1, ..., p_{37})^T$ based on observing $\hat{\boldsymbol{p}} = (\hat{p}_1, ..., \hat{p}_{37})^T$. In this problem, we use data from the dataset DirectEstimation.txt, which contains values for $\operatorname{logit}(\hat{\boldsymbol{p}})$ and standard deviations $\sqrt{V_a}$, $a = 1, ..., 37$ as well as the names of the admin1 areas. We consider hierarchical spatial models where we imagine the true proportions to be stochastic variables. We define the notation $\boldsymbol{X} = (\operatorname{logit}(P_1), ..., \operatorname{logit}(P_{37}))^T$ and $\boldsymbol{Y} = (\operatorname{logit}(\hat{P}_1), ..., \operatorname{logit}(\hat{P}_{37}))^T$ for the random vector of true proportions and the random vector of observed proportions respectively.

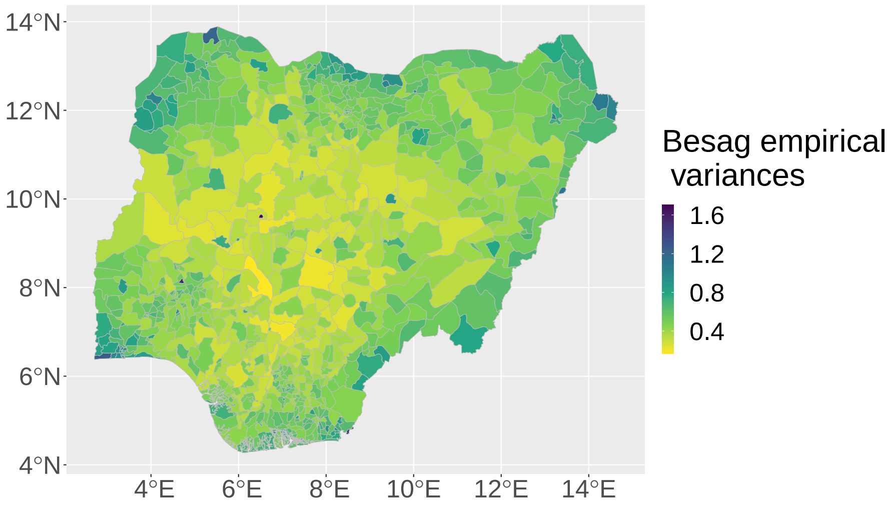


Figure 4: Empirical variances for admin2 Besag model

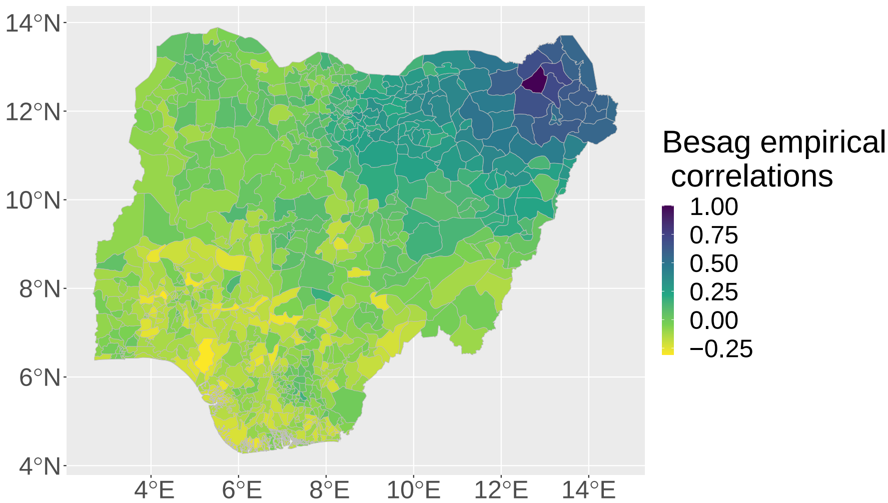


Figure 5: Empirical correlations for Gubio in admin2

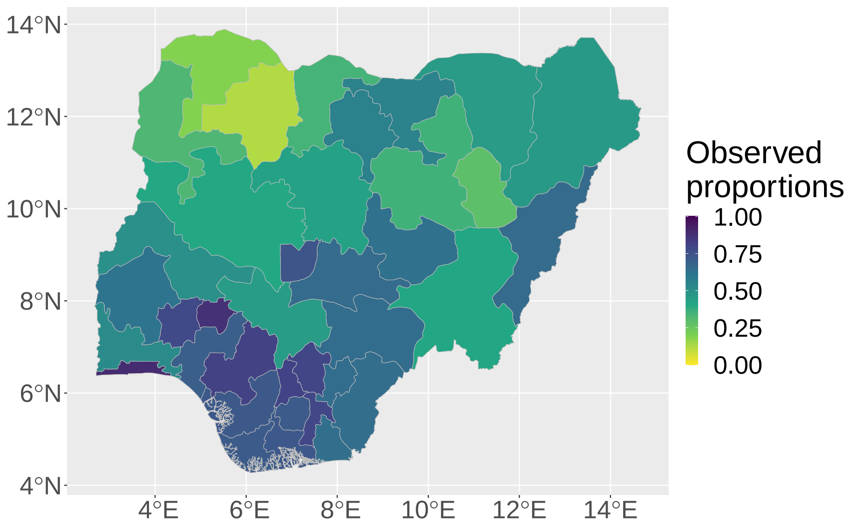


Figure 6: Observed proportions ($\hat{\boldsymbol{p}}$) admin1

3.1 a)

In this section we will look at the observed proportions from the admin1 administrative division, and determine whether a spatial model seems reasonable. In Figure 6 we can see that there is a spatial structure in the estimated vaccination rates. Areas in the North-West of Nigeria seem to be less vaccinated than the South, and we can assume that this is a spatial trend. This may be because this area is poorer, or has worse living conditions than the South of the country. Thus using a spatial model to reduce uncertainty seems like a reasonable choice.

3.2 b)

From the assumption in equation (5), we find the distribution

$$\boldsymbol{Y}|\boldsymbol{X} = \boldsymbol{x} \sim \mathcal{N}_{37}(\boldsymbol{x}, \mathbf{D}) \tag{6}$$

where $\mathbf{D}$ is a 37×37 diagonal matrix with diagonal elements $d_{a,a} = V_a$, $a = 1, ..., 37$. Now we assume the vague prior $\boldsymbol{X} \sim \mathcal{N}_{37}(\mathbf{0}, \sigma^2\mathbf{I})$, where $\sigma^2 = 100^2$. We want to determine the distribution $\boldsymbol{X}|\boldsymbol{Y} = \boldsymbol{y}$ and its parameters. We find that

$$\begin{aligned} f(\boldsymbol{x}, \boldsymbol{y}) &= f(\boldsymbol{x})f(\boldsymbol{y}|\boldsymbol{x}) && (7)\\ &\propto \exp\left(-\frac{1}{2}\sigma^{-2}\boldsymbol{x}^T\boldsymbol{x}\right)\exp\left(-\frac{1}{2}(\boldsymbol{y}-\boldsymbol{x})^T\mathbf{D}^{-1}(\boldsymbol{y}-\boldsymbol{x})\right) && (8)\\ &= \exp\left(-\frac{1}{2}\boldsymbol{x}^T(\sigma^{-2}\mathbf{I}+\mathbf{D}^{-1})\boldsymbol{x} + \boldsymbol{x}^T\mathbf{D}^{-1}\boldsymbol{y} - \frac{1}{2}\boldsymbol{y}^T\mathbf{D}^{-1}\boldsymbol{y}\right) && (9) \end{aligned}$$

To find the distribution $\boldsymbol{X}|\boldsymbol{Y} = \boldsymbol{y}$ and its parameters, we use the fact that it is a GMRF on admin1, and we know that we can write it on form

$$\begin{aligned} f(\boldsymbol{x}, \boldsymbol{y}) &\propto \exp\left(-\frac{1}{2}(\boldsymbol{x}, \boldsymbol{y})\begin{pmatrix}\mathbf{Q}_{xx} & \mathbf{Q}_{xy}\\ \mathbf{Q}_{yx} & \mathbf{Q}_{yy}\end{pmatrix}\begin{pmatrix}\boldsymbol{x}\\ \boldsymbol{y}\end{pmatrix}\right) && (10)\\ &= \exp\left(-\frac{1}{2}\boldsymbol{x}^T\mathbf{Q}_{xx}\boldsymbol{x} - \boldsymbol{x}^T\mathbf{Q}_{xy}\boldsymbol{y} - \frac{1}{2}\boldsymbol{y}^T\mathbf{Q}_{yy}\boldsymbol{y}\right) && (11) \end{aligned}$$

Equation the the terms in equation (11) and equation (9), we find that $\mathbf{Q}_{xx} = \sigma^{-2}\mathbf{I} + \mathbf{D}^{-1}$, $\mathbf{Q}_{xy} = \mathbf{Q}_{yx} = -\mathbf{D}^{-1}$ and $\mathbf{Q}_{yy} = \mathbf{D}^{-1}$. Then the distribution of $\boldsymbol{X}|\boldsymbol{Y} = \boldsymbol{y}$ is given by

$$\boldsymbol{X}|\boldsymbol{Y}=\boldsymbol{y}\sim\mathcal{N}_{37}(\boldsymbol{\mu}_{\boldsymbol{x}|\boldsymbol{y}},\mathbf{Q}_{\boldsymbol{x}|\boldsymbol{y}}^{-1})$$

where

$$\begin{aligned}\boldsymbol{\mu}_{\boldsymbol{x}|\boldsymbol{y}}=-\mathbf{Q}_{xx}^{-1}\mathbf{Q}_{xy}\boldsymbol{y}&=(\sigma^{-2}\mathbf{I}+\mathbf{D}^{-1})^{-1}\mathbf{D}^{-1}\boldsymbol{y}\\&=(\sigma^{-2}\mathbf{D}+\mathbf{I})^{-1}\boldsymbol{y}\end{aligned}$$

and

$$\mathbf{Q}_{\boldsymbol{x}|\boldsymbol{y}}=\mathbf{Q}_{xx}=\sigma^{-2}\mathbf{I}+\mathbf{D}^{-1}$$

We see that in the limiting case where $\sigma^2\to\infty$, we have that

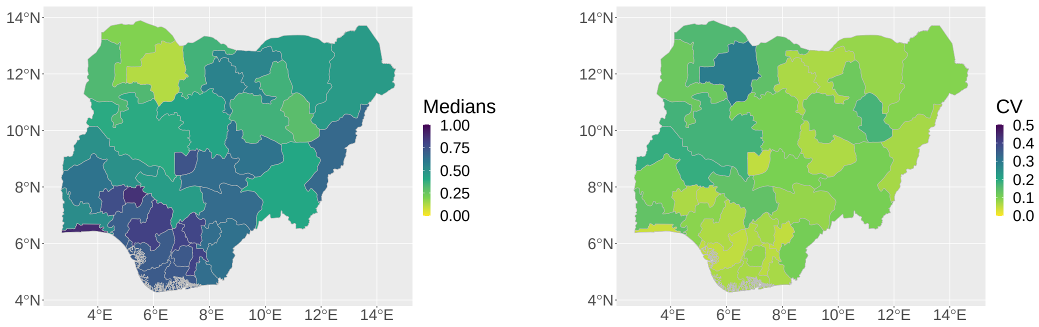
$$\lim_{\sigma^2\to\infty}\boldsymbol{X}|\boldsymbol{Y}=\boldsymbol{y}\sim\mathcal{N}_{37}(\boldsymbol{y},\mathbf{D})$$

From the assumption given in equation (5), we find that in this limiting case, $P_a|\boldsymbol{Y}=\boldsymbol{y}$ and its parameters for $a=1,...,37$ has distribution

$$P_a|Y_a=y_a\sim\text{logitNormal}(p_a;y_a,V_a)$$

with probability density function

$$f(p_a;y_a,V_a)=\frac{1}{\sqrt{V_a2\pi}}\frac{1}{p_a(1-p_a)}\exp\left(-\frac{1}{2}\frac{(\text{logit}(p_a)-y)^2}{V_a}\right)$$



(a) Estimated medians

(b) Estimated CV

Figure 7: Estimated CV and medians using non-Besag hierarchical Bayesian model in admin1

The simulated values $P_a|\boldsymbol{Y}=\boldsymbol{y}$ from Figure 7 using the non-Besag hierachical Bayesian model gave estimates which are very close the the estimates given in Figure 6. Even though we cannot see from the expressions above that the median is exactly equal to $\boldsymbol{y}$, we see that with $\sigma^2=100^2$ they are essentially indistinguishable. We also see that the coefficient of variation (CV) is greater in the peripheral areas in Nigeria, indicating that there is more uncertainty in the estimates there.

3.3 c)

Now we will assume that $\boldsymbol{X}$ a priori follows a Besag model, defined on the admin1 area graph, with precision parameter τ. We want to determine the expected value and precision matrix of $\boldsymbol{X}|\boldsymbol{Y}=\boldsymbol{y}$ as a function of τ in this case. We find that

$$
\begin{aligned}
f(\boldsymbol{x},\boldsymbol{y}) &= f(\boldsymbol{x})f(\boldsymbol{y}|\boldsymbol{x})\\
&\propto \exp\left(-\frac{\tau}{2}\sum_{i\sim j}(x_i-x_j)^2\right)\exp\left(-\frac{1}{2}(\boldsymbol{y}-\boldsymbol{x})^T\mathbf{D}^{-1}(\boldsymbol{y}-\boldsymbol{x})\right)\\
&= \exp\left(-\frac{\tau}{2}\sum_{i\sim j}(x_i^2-2x_ix_j+x_j^2)\right)\exp\left(-\frac{1}{2}(\boldsymbol{y}^T\mathbf{D}^{-1}\boldsymbol{y}-2\boldsymbol{x}^T\mathbf{D}^{-1}\boldsymbol{y}+\boldsymbol{x}^T\mathbf{D}^{-1}\boldsymbol{x}\right)
\end{aligned}
$$

Equating again the terms in this equation with the terms in equation (11), we find that $\mathbf{Q}_{yy}=\mathbf{D}^{-1}$, $\mathbf{Q}_{xy}=\mathbf{Q}_{yx}=-\mathbf{D}^{-1}$ and $\mathbf{Q}_{xx}=\tau\mathbf{R}_1+\mathbf{D}^{-1}$, where we use $\tau\mathbf{R}_1$ from equation (4). Then we find that

$$\boldsymbol{\mu}_{\boldsymbol{x}|\boldsymbol{y}} = -\mathbf{Q}_{xx}\mathbf{Q}_{xy}\boldsymbol{y} = (\tau\mathbf{R}_1+\mathbf{D}^{-1})^{-1}\mathbf{D}^{-1}\boldsymbol{y}$$

and

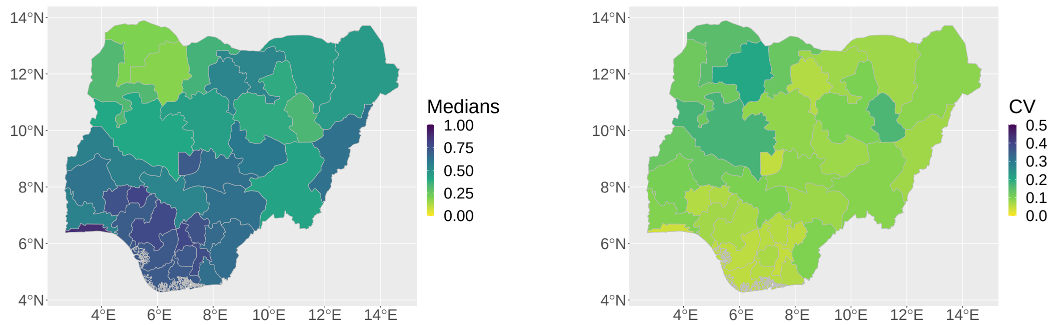
$$\mathbf{Q}_{\boldsymbol{x}|\boldsymbol{y}} = \mathbf{Q}_{xx} = \tau\mathbf{R}_1+\mathbf{D}^{-1}$$

and we have the distribution

$$\boldsymbol{X}|\boldsymbol{Y}=\boldsymbol{y}\sim\mathcal{N}(\boldsymbol{\mu}_{\boldsymbol{x}|\boldsymbol{y}},\mathbf{Q}_{\boldsymbol{x}|\boldsymbol{y}}^{-1}) \tag{12}$$

Since the matrix $\mathbf{Q}_{\boldsymbol{x}|\boldsymbol{y}}$ is a linear combination of $\mathbf{D}^{-1}$, and $\mathbf{D}$ is full rank (and thus $\mathbf{D}^{-1}$ also), $\mathbf{Q}_{\boldsymbol{x}|\boldsymbol{y}}$ is also full rank, and thus $\boldsymbol{X}|\boldsymbol{Y}=\boldsymbol{y}$ is a proper GMRF.

Simularly to Section b) we will generate 100 realizations from $P_a|\boldsymbol{Y}=\boldsymbol{y}$, but this time using the Besag model. We see that using the Besag model, we see more spatial "smoothing" between



(a) Estimated medians

(b) Estimated CV

Figure 8: Estimated CV and medians using Besag hierarchical model in admin1

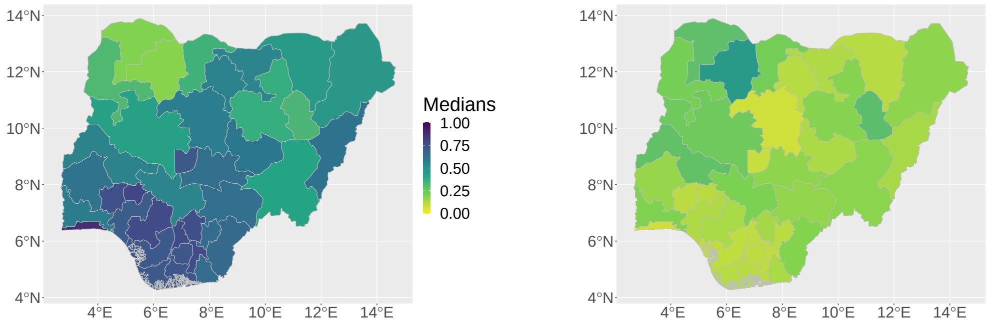
neighbouring areas in admin1. This is expected, since we assume that neighbouring states are correlated in the Besag model. We also see a reduction in CV, as we borrow strength from the spatial structure, such that the CV in the North-West of Nigeria for example is reduced significantly.

3.4 d)

In this section we work under the assumption that in addition to the national survey that gave rise to $\boldsymbol{Y}$, 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 $Y_{38}|\boldsymbol{P}$ is independent of $\boldsymbol{Y}|\boldsymbol{P}$. We want to produce better estimates for the vaccination coverages using all 38 observations. We let $\tilde{\boldsymbol{Y}}=(Y_1,...,Y_{37},Y_{38})^T$ and $\tilde{\mathbf{D}}\in\mathbb{R}^{38\times38}$ be the extension of $\mathbf{D}$ such that the last row is given by $(0,...,0.1^2)$. We assume that $\boldsymbol{X}$ follows the same Besag model as in Section c). Now we want to determine the expected value and precision matrix of $\boldsymbol{X}|\tilde{\boldsymbol{Y}}=\tilde{\boldsymbol{y}}$ as a function of τ as previously done. We define the matrix $\mathbf{M}\in\mathbb{R}^{38\times37}$ such that the



(a) Estimated medians

(b) Estimated CV

Figure 9: Estimates of CV and medians using Besag on admin1 and Kaduna estimate

final row of $\mathbf{M}$ is the vector $(0, ..., 1, ..., 0)^T$ with the element 1 located at the index corresponding to Kaduna in the vector $\boldsymbol{X}$ (index 19). Further, rows and columns 1 through 37 constitute the 37×37 identity matrix. We then have the distribution

$$\tilde{\boldsymbol{Y}} | \boldsymbol{X} = \boldsymbol{x} \sim \mathcal{N}_{38}(\mathbf{M}\boldsymbol{x}, \tilde{\mathbf{D}})$$

We again find that

$$\begin{aligned} f(\boldsymbol{x}, \boldsymbol{y}) &= f(\boldsymbol{x}) f(\boldsymbol{y}|\boldsymbol{x}) \\ &\propto \exp\left(-\frac{\tau}{2}\sum_{i\sim j}(x_i^2 - 2x_i x_j + x_j^2)^2\right) \exp\left(-\frac{1}{2}(\boldsymbol{y} - \mathbf{M}\boldsymbol{x})^T \tilde{\mathbf{D}}^{-1}(\boldsymbol{y} - \mathbf{M}\boldsymbol{x})\right) \\ &= \exp\left(-\frac{\tau}{2}\sum_{i\sim j}(x_i^2 - 2x_i x_j + x_j^2) - \frac{1}{2}(\boldsymbol{y}^T \tilde{\mathbf{D}}^{-1}\boldsymbol{y} - 2\boldsymbol{x}^T \mathbf{M}^T \tilde{\mathbf{D}}^{-1}\boldsymbol{y} + \boldsymbol{x}^T \mathbf{M}^T \tilde{\mathbf{D}}^{-1}\mathbf{M}\boldsymbol{x})\right) \end{aligned}$$

We again equate the terms in this equation with the terms in equation (11) and we find that $\mathbf{Q}_{yy} = \tilde{\mathbf{D}}^{-1}$, $\mathbf{Q}_{xy} = \mathbf{Q}_{yx} = -\mathbf{M}^T \tilde{\mathbf{D}}^{-1}$ and $\mathbf{Q}_{xx} = \tau \mathbf{R}_1 + \mathbf{M}^T \tilde{\mathbf{D}}^{-1}\mathbf{M}$. We then have that

$$\boldsymbol{X} | \tilde{\boldsymbol{Y}} = \tilde{\boldsymbol{y}} \sim \mathcal{N}_{37}(\boldsymbol{\mu}_c, \mathbf{Q}_c^{-1}) \tag{13}$$

where

$$\begin{aligned} \boldsymbol{\mu}_c := \boldsymbol{\mu}_{\boldsymbol{x}|\tilde{\boldsymbol{y}}} &= -\mathbf{Q}_{xx}^{-1}\mathbf{Q}_{xy}\tilde{\boldsymbol{y}} \\ &= (\tau \mathbf{R}_1 + \mathbf{M}^T \tilde{\mathbf{D}}^{-1}\mathbf{M})^{-1}\mathbf{M}^T \tilde{\mathbf{D}}^{-1}\tilde{\boldsymbol{y}} \end{aligned}$$

and

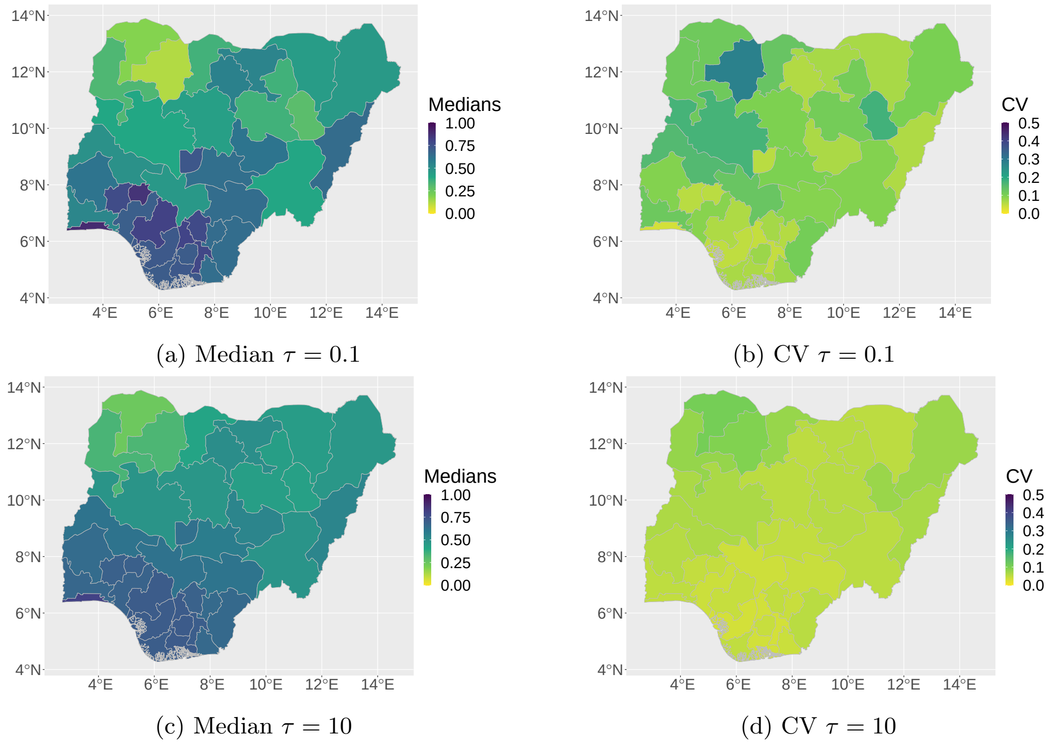
$$\begin{aligned} \mathbf{Q}_c := \mathbf{Q}_{\boldsymbol{x}|\tilde{\boldsymbol{y}}} &= \mathbf{Q}_{xx} \\ &= \tau \mathbf{R}_1 + \mathbf{M}^T \tilde{\mathbf{D}}^{-1}\mathbf{M} \end{aligned}$$

Similarly to the previous sections we will generate 100 realizations from the model, and compare how the additional survey information affects the estimates.

From Figure 9 we see that adding information on Kaduna greatly affects the estimate of one of the states in the admin1 administrative division, and thus also affects the estimates of the neighbouring states. From the right side of the figure, we see that the CV of Kaduna is much smaller and close to zero. This is to be expected since we add an additional estimate in this area with variance 0.1^2. For areas other than Kaduna, it looks like the CV's are unchanged or very close to unchanged.

3.5 e)

In this section we want to compare realizations using different values for τ, namely $\tau = 0.1$, $\tau = 1$ and $\tau = 10$. We want to see how the precision parameter affects the realizations, and how sensitive



(a) Median $\tau = 0.1$
(b) CV $\tau = 0.1$
(c) Median $\tau = 10$
(d) CV $\tau = 10$

Figure 10: Estimates median and CV using Besag model on admin1 for different values of τ

they are to changes in the parameter. When looking at Figure 10 and 8 we see that changing the values of τ indeed affects the generated realizations. τ acts as a smoothing parameter, where increasing τ increases similarities between neighbouring states. Based on these Figures, we can conclude that estimating τ correctly is indeed important for our model.

3.6 f)

In this section we want to estimate τ by maximum likelihood estimation. We start by applying Bayes theorem and we find that

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

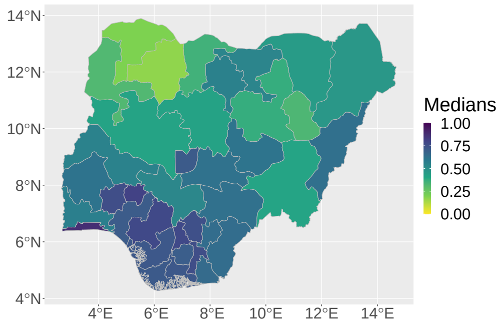
Taking the logarithm of this equation, we find the log-likelihood function $l(\tau;\boldsymbol{y})$:

$$l(\tau;\boldsymbol{y}) = \log f(\boldsymbol{y};\tau) = \log f(\boldsymbol{x};\tau) + \log f(\boldsymbol{y}|\boldsymbol{x}) - \log f(\boldsymbol{x}|\boldsymbol{y};\tau)$$

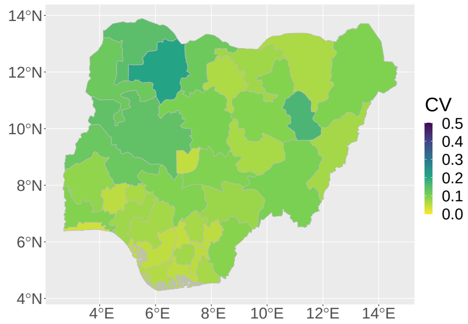
Still assuming that $\boldsymbol{X}$ follows the Besag model with probability density function given in equation (1), that $\boldsymbol{Y}|\boldsymbol{X} = \boldsymbol{x} \sim \mathcal{N}_{37}(\boldsymbol{x}, \mathbf{D})$ and that $\boldsymbol{X}|\boldsymbol{Y} = \boldsymbol{y} \sim \mathcal{N}_{37}(\boldsymbol{\mu}_c, \mathbf{Q}_c^{-1})$, we find the following expression for the log-likelihood function:

$$\begin{aligned} l(\tau;\boldsymbol{y}) = \text{Const} + \frac{37-1}{2}\log(\tau) - \frac{\tau}{2}\boldsymbol{x}^T\mathbf{R}_1\boldsymbol{x} - \frac{1}{2}(\boldsymbol{y}-\boldsymbol{x})^T\mathbf{D}^{-1}(\boldsymbol{y}-\boldsymbol{x}) \\ - \frac{1}{2}\log|\mathbf{Q}_c| + \frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_c)^T\mathbf{Q}_c(\boldsymbol{x}-\boldsymbol{\mu}_c) \end{aligned} \tag{14}$$

Using the R-function `optimize()` with Eq. (14) we can obtain an estimate for the precision parameter. As we do not have the value for $\boldsymbol{x}$ in this equation, we set it to $\mathbf{0}$. We note that solving the optimization problem with different reasonable initializations of $\boldsymbol{x}$ gives almost identical results. Doing this, we found the estimate to be $\hat{\tau} \approx 0.806$, which is quite close to the inital value of 1. Having computed the maximum likelihood estimate, we again generate 100 realizations of the Besag model, and compare the median and CV with the previous sections.



(a) Median $\tau = \hat{\tau} = 0.806$



(b) CV $\tau = \hat{\tau} = 0.806$

Figure 11: Besag model estimates with maximum likelihood precision parameter

As the maximum likelihood estimate is quite close the the initial value $\tau = 1$, we do not see differences that are obvious between this model and the model we used in Section c).

Bibliography

BATH, University of (2024). *INLA for Spatial Statistics*. URL: https://faculty.washington.edu/jonno/SISMIDmaterial/3-spatial1.pdf (visited on 24th Apr. 2024).