Structure induced by a multiple membership transformation on the Conditional Autoregressive model

Marco Gramatica, Peter Congdon, Silvia Liverani

Queen Mary University of London

The "usual" GLM

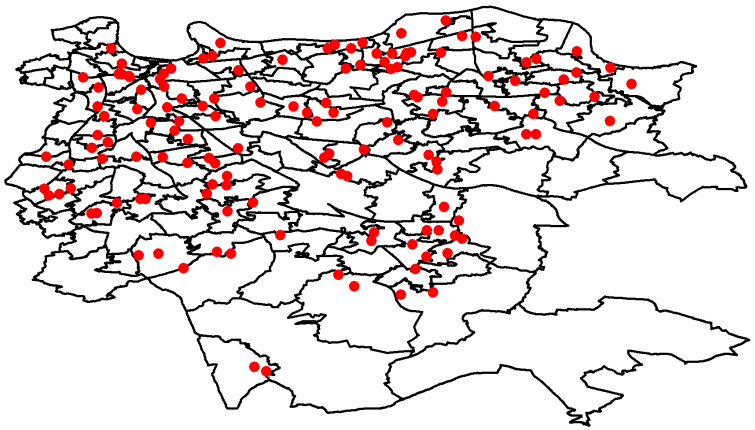
Observed outcome

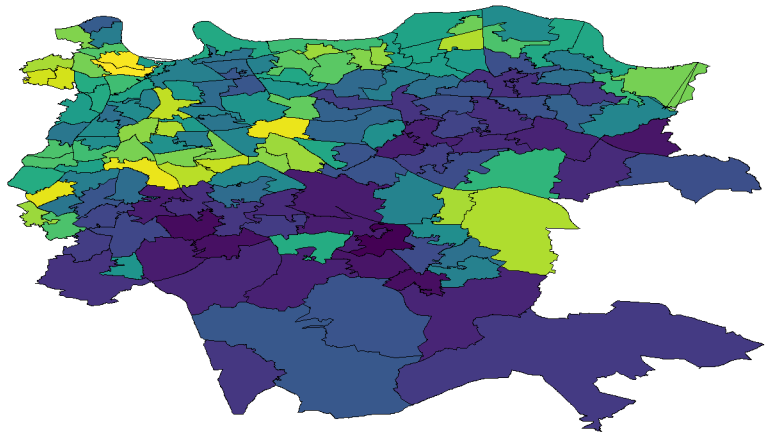
$\tilde{Y}_j \sim Poi(\tilde{E}_j \tilde{\rho}_j)\,;\, j = 1, ..., m$

Prevalence of a disease by *General Practitioner* (GP)

Covariates

- Ethnic makeup of the population
- Pollution
- Deprivation :





Spatial Misalignment

Misaligned relative risk

- We have the expected rates $\tilde{E}_j$ by GP (memberships)
- The remaining part is the relative risk:

$$\log \tilde{\rho}_j = g(\gamma + x_i^T \boldsymbol{\beta} + \phi_i) \qquad i = 1, ..., n$$

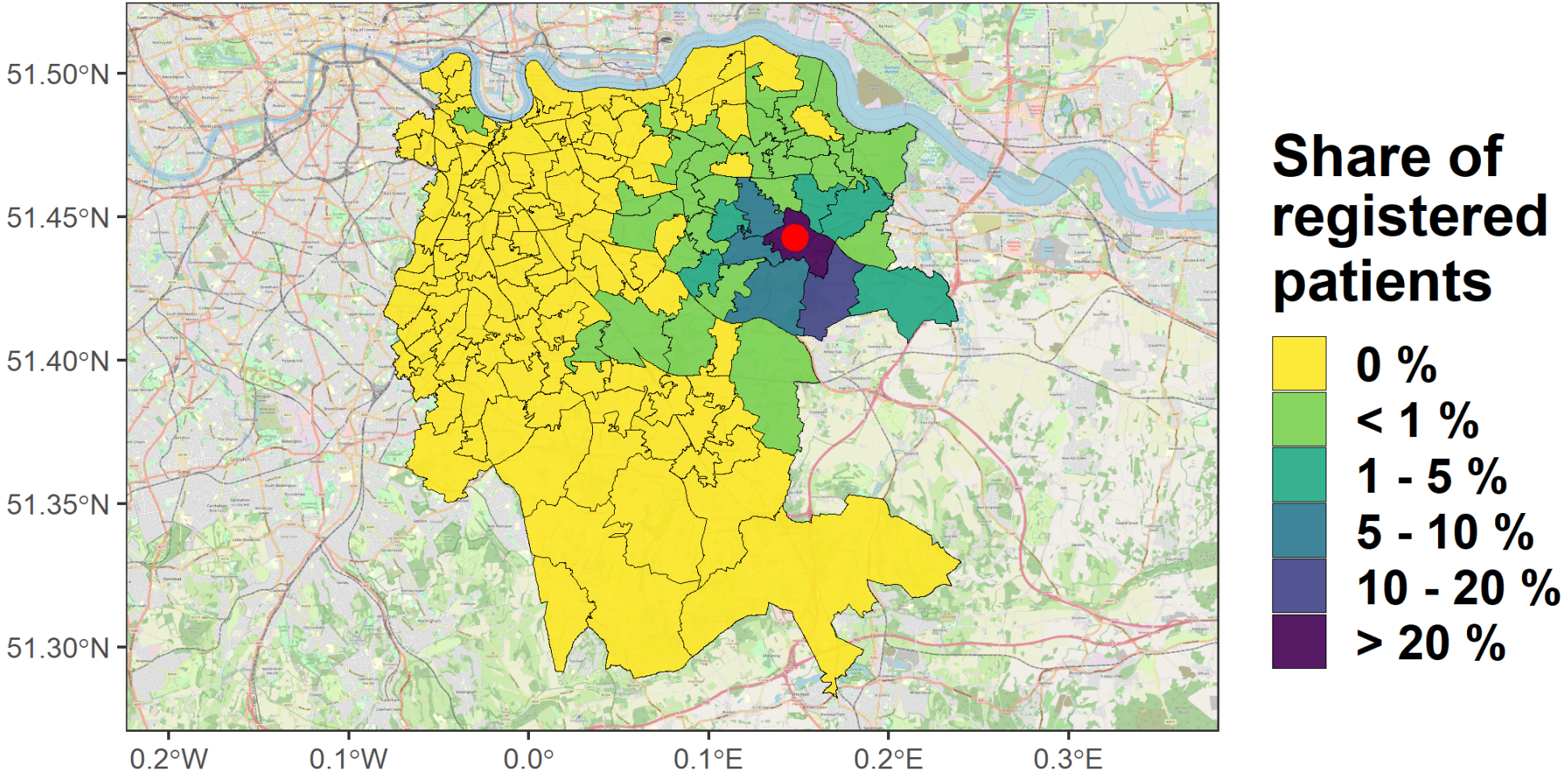
$$\boldsymbol{\phi} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}^{-1} = \tau(\boldsymbol{D} - \alpha \boldsymbol{W})) \quad (CAR)$$

- $\boldsymbol{D}$: is diagonal, contains the number of neighbours of each area
- $\boldsymbol{W}$: is the adjacency matrix
- $\alpha \in [0, 1)$, $\tau \in \mathbb{R}^+$

How do we choose $g(\cdot)$?

Idea: share of patients of practice j, that reside in area i

How do we choose $g(\cdot)$?



How do we choose $g(\cdot)$? $\boldsymbol{H}$

- $\boldsymbol{H}$: $(m \times n)$ matrix, (membership $\times$ area)
- Each row has the share of patients of a practice that reside in a certain area : **Multiple Membership principle (MM)**
- Use **H** to *average* areal relative risks into membership relative risks
- Requirements for **H**:
  - for each row $j = 1, ..., m$ of **H**: $\sum_{i=1}^{n} (H)_{ji} = 1$, $0 \leq (H)_{ji} \leq 1$ for all i and j
  - it is of full rank
  - there is no column i such that $(\mathbf{H})_{\cdot i} = \mathbf{0}$

$$\tilde{\boldsymbol{Y}} \sim Poi(\tilde{\boldsymbol{E}} \circ \tilde{\boldsymbol{\rho}})$$

$$\boxed{\log \tilde{\boldsymbol{\rho}} = \log \boldsymbol{H}\boldsymbol{\rho} = \log \boldsymbol{H}(\gamma + X\boldsymbol{\beta} + \boldsymbol{\phi})}$$

A new prior

Since $\mathbf{H}$ is a *linear* transformation and ϕ *Gaussian*:

$$\mathbf{H}\phi = \tilde{\phi} \sim \mathcal{N}_m(\mathbf{0}, \mathbf{H}\boldsymbol{\Sigma}\mathbf{H}^T = \tilde{\boldsymbol{\Sigma}})$$

$\tilde{\phi}$ is a new CAR prior we will call: **CAR-MM**.

We have two possible parameterisations:

- *Post:* $\phi \sim \mathcal{N}_n(\mathbf{0}, \boldsymbol{\Sigma}) \Rightarrow \tilde{\phi} = \mathbf{H}\phi$

  Sample from the areal random effects
- *Inverse:* $\tilde{\phi} \sim \mathcal{N}_m(\mathbf{0}, \tilde{\boldsymbol{\Sigma}}) \Rightarrow \phi = \mathbf{H}^{-}\tilde{\phi}$

  Sample from the membership random effects

Implementation of the CAR prior : **RStan**

$\boldsymbol{H}$: ($m \times n$) matrix, (membership $\times$ area)

```
model{

   phi ~ sparse_car(
         alpha, W_sparse, D_sparse, lambda, n, W_n
         );

         // [...] Priors for the other parameters [...]

   y ~ poisson(exp(log_offset + r_1));

}
transformed parameters{
         vector[m] r_1; // Memberships relative risks
         r_1 = H*(gamma + X_cov * beta + phi);
}
```

Implementation of the CAR prior : **RStan**

CAR precision matrix : $\mathbf{Q} = \boldsymbol{\Sigma}^{-1} = \tau(\boldsymbol{D} - \alpha \boldsymbol{W})$

- $\boldsymbol{D}$ is *diagonal*, contains the number of neighbours of each area: n_i
- $\boldsymbol{W}$ is the adjacency matrix. Contains only 0 and 1's
- $\alpha \in [0, 1)$, $\tau \in \mathbb{R}^+$

Problem: $n \times n$ matrix can take too long to evaluate even for moderate n.

Solution: make use of 3 tricks to compute the Gaussian density quickly

Implementation of the CAR prior : **RStan**

1. **Gaussian prior sparsity** (Joseph, 2016)
2. **Fast determinant computation** (Jin et al. 2005)
3. **Non-centred parameterisation**

- The first two approaches can only be used for the **post** parameterisation
- The **inverse** parameterisation can be implemented with the `multi_normal_prec` function

1. Gaussian prior sparsity (Joseph, 2016)

$$\phi^T \mathbf{Q} \phi = \sum_{i=1}^{n} \sum_{j=1}^{n} \phi_i \phi_j (\mathbf{Q})_{ij} \qquad (\mathbf{Q})_{ij} = 0 \text{ iff } i \not\sim j$$

`W_sparse`: $(\sum_{i=1}^{n} n_i) \times 2$ matrix encoding all the "adjacencies"

```
row_vector[n] phit_D; // Diagonal elements
row_vector[n] phit_W; // Adjacent nodes (or areas)
vector[n] ldet_terms; // Determinant

phit_D = (phi .* D_sparse)';
phit_W = rep_row_vector(0, n);
for (i in 1:W_n) { // W_n : no of adjacent node pairs
    phit_W[W_sparse[i, 1]] =
              phit_W[W_sparse[i, 1]] + phi[W_sparse[i, 2]];
    phit_W[W_sparse[i, 2]] =
              phit_W[W_sparse[i, 2]] + phi[W_sparse[i, 1]];
}
```

2. Fast determinant computation (Jin et al. 2005)

$$\log(\det((\mathbf{D} - \alpha\mathbf{W}))) = \log(\det(\mathbf{D})) + \sum_{i=1}^{n}(1 - \alpha\lambda_i),$$

```
for (i in 1:n) ldet_terms[i] = log1m(alpha * lambda[i]);
```

3. Non-centred parameterisation

```
transformed parameters{
        vector[m] r_1; // Relative risk
        real<lower = 0> invtausq;
        vector[n] phi_unscaled;

        invtausq = inv_sqrt(tau);
        phi = invtausq*phi_unscaled;

        r_1 = H*(gamma + X_cov * beta + phi);
}
```

Identifiability - Post parameterisation

Definition (Posterior non-identifiability [3]):
A Bayesian model $M(\boldsymbol{\psi}, \boldsymbol{y})$ described by the posterior distribution, is globally identifiable if $f(\boldsymbol{\psi}_a \mid \boldsymbol{y}) = f(\boldsymbol{\psi}_b \mid \boldsymbol{y})$ implies $\boldsymbol{\psi}_a = \boldsymbol{\psi}_b$.

$$\boldsymbol{\psi} = (\boldsymbol{\phi}, \gamma, \boldsymbol{\beta}) = (\boldsymbol{\phi}, \boldsymbol{\theta}) \quad ; \quad \boldsymbol{\Delta} = \{\tilde{\boldsymbol{y}}, \mathbf{H}, \boldsymbol{X}\}$$

- We take $g(\cdot) \equiv \mathbf{H}$, which is deterministic
- *Iff* $m \geq n$, $\mathbf{H}$ is injective (more memberships than areas)

Post parameterisation:

$$f(g(\boldsymbol{\phi}_a) \mid \boldsymbol{\phi}_a, \boldsymbol{\theta}_a, \boldsymbol{\Delta}) f(\boldsymbol{\phi}_a, \boldsymbol{\theta}_a \mid \boldsymbol{\Delta}) = f(g(\boldsymbol{\phi}_b) \mid \boldsymbol{\phi}_b, \boldsymbol{\theta}_b, \boldsymbol{\Delta}) f(\boldsymbol{\phi}_b, \boldsymbol{\theta}_b \mid \boldsymbol{\Delta})$$
$$f(\boldsymbol{\phi}_a, \boldsymbol{\theta}_a \mid \boldsymbol{\Delta}) = f(\boldsymbol{\phi}_b, \boldsymbol{\theta}_b \mid \boldsymbol{\Delta}).$$

The *post* parameterised CAR-MM is identifiable iff $m \geq n$

Identifiability - Inverse parameterisation

$$\psi = (\phi, \gamma, \beta) = (\phi, \theta) \quad ; \quad \Delta = \{\tilde{y}, \mathbf{H}, X\}$$

- We take $k(\cdot) \equiv \mathbf{H}^{-}$, which is deterministic
- The matrix $\mathbf{H}^{-}$:
  - exists *iff* $m \geq n$
  - is injective *iff* $m = n$

Inverse parameterisation:

$$f(k(\tilde{\phi}_a) \mid \tilde{\phi}_a, \theta_a, \Delta) f(\tilde{\phi}_a, \theta_a \mid \Delta) = f(k(\tilde{\phi}_b) \mid \tilde{\phi}_b, \theta_b, \Delta) f(\tilde{\phi}_b, \theta_b \mid \Delta)$$
$$f(\tilde{\phi}_a, \theta_a \mid \Delta) = f(\tilde{\phi}_b, \theta_b \mid \Delta).$$

The *Inverse* parameterised CAR-MM is identifiable iff $m = n$

Post or *Inverse*?

Inverse parameterisation precision matrix specification:

$$\tilde{\boldsymbol{Q}} = \left(\mathbf{H}\mathbf{Q}^{-1}\mathbf{H}^{T}\right)^{-1}$$

1. *Post* is identifiable for $m \geq n$, instead of just $m = n$
2. Computationally, the *Post* parameterisation can rely on sparse representations, given that we sample from a CAR prior and *post*-transform the random effects

Post parameterisation is preferable

Simulation Based Calibration (**SBC**) (Talts et al., 2020) [4]

$$\pi(\boldsymbol{\theta}, \tilde{\boldsymbol{y}}) = l(\tilde{\boldsymbol{y}}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}) \propto \pi(\boldsymbol{\theta}|\tilde{\boldsymbol{y}})$$

$$\pi(\boldsymbol{\theta}) = \int \pi(\boldsymbol{\theta}|\tilde{\boldsymbol{y}})\pi(\boldsymbol{\theta}, \tilde{\boldsymbol{y}}) d\tilde{\boldsymbol{\theta}} d\tilde{\boldsymbol{y}}$$

Posterior samples should be distributed like the prior

Simulation Based Calibration (**SBC**) (Talts et al., 2020) [4]

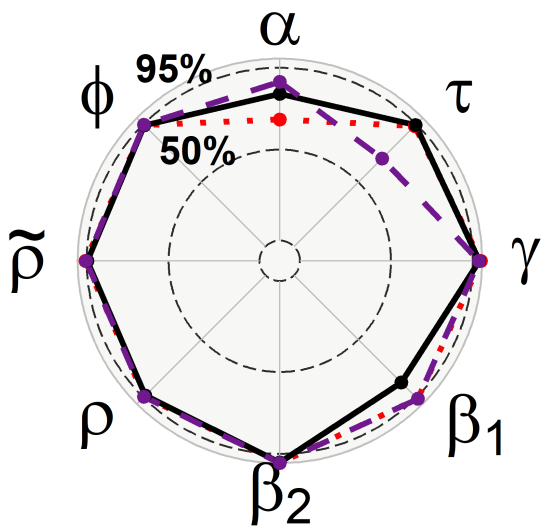
We simulate 10^4 dataset:

- Spatial structure: 10×10 grid ($n = 100$)
- Three membership sizes: $m = 70, 100, 130$
- We evaluate both parameterisations for $m = 70, 100$, but for $m = 130$ only the *Post*
- $\gamma, \beta_1, \beta_2 \sim \mathcal{N}(0, 0.49)$
- $\alpha \sim UC(0, 1)$
- $\tau \sim Gamma(2, 0.2)$
- Covariate matrix $\boldsymbol{X}$ from independent standard Gaussians min-max normalised
- Outcomes: $\tilde{Y}_j \sim Poisson(\tilde{E}_j \tilde{\rho}_j)$

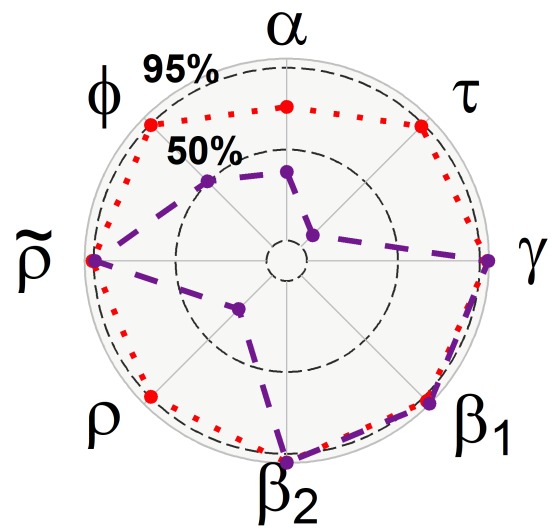
For each generated dataset, we assess calibration between "true" parameters and posterior samples.
All models were coded in **Rstan** [5] [6].

Calibration study: coverage of expected variation

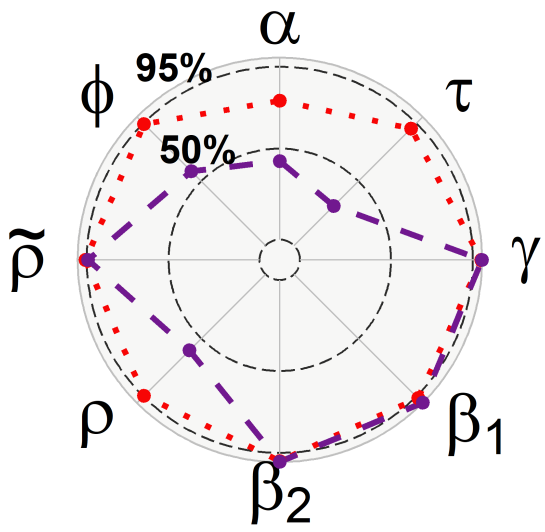
Data: Post - MCMC: Post



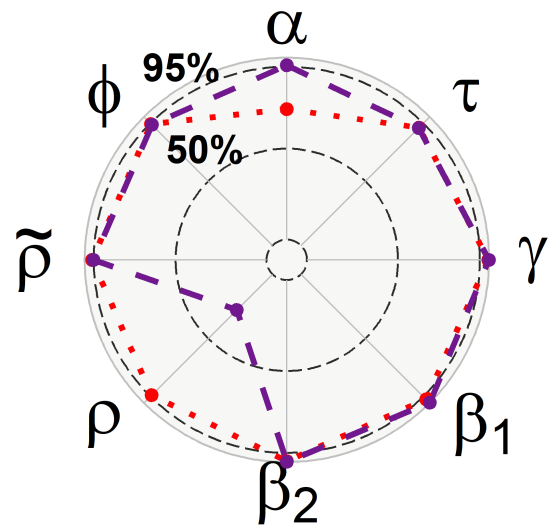
Data: Post - MCMC: Inverse



Data: Inverse - MCMC: Post



Data: Inverse - MCMC: Inverse



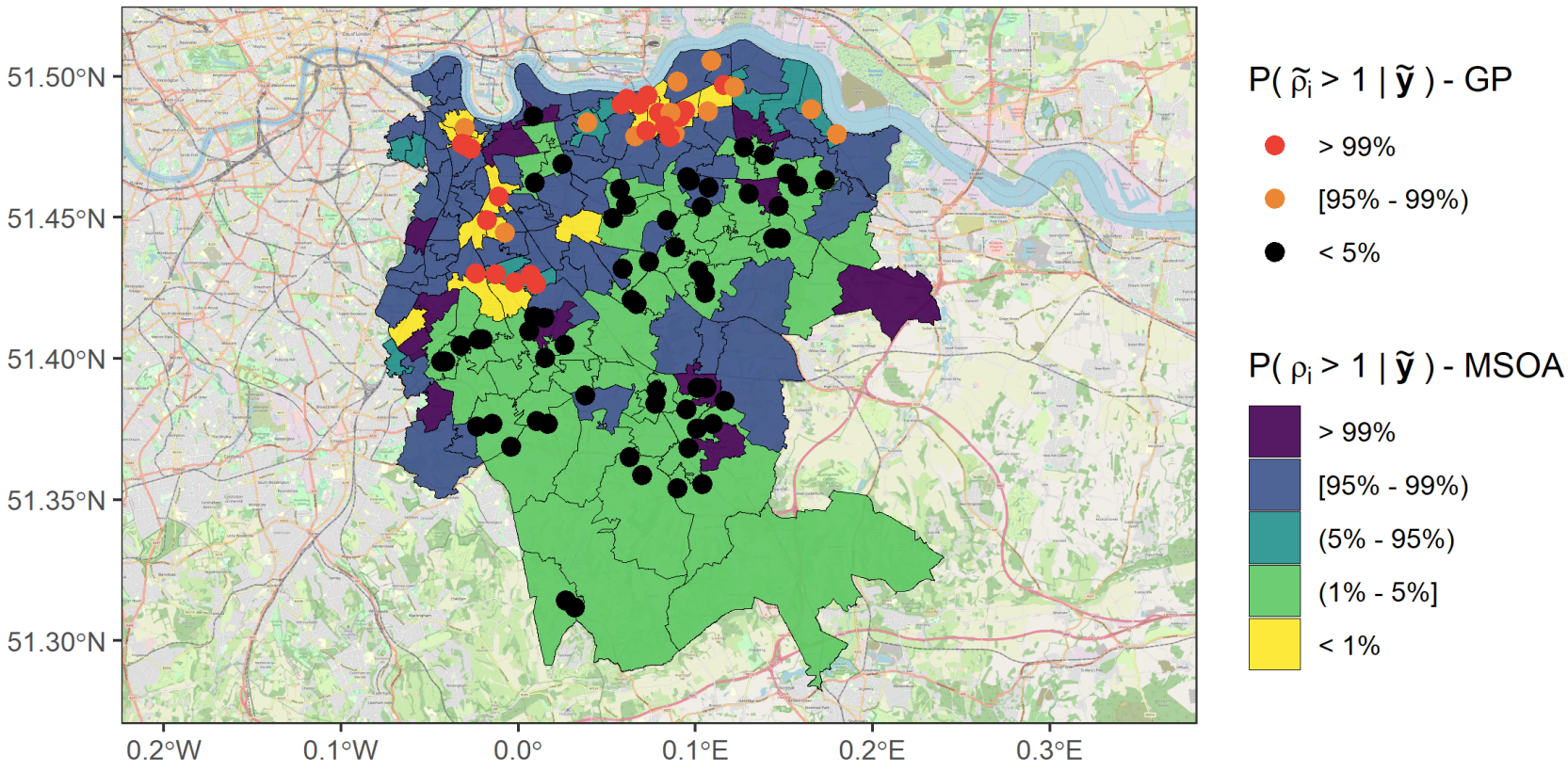
Number of memberships — 130 ··· 100 -- 70

Data Analysis

We analyse a dataset of patients registered with diabetes in South East London:

- Index of multiple deprivation
- Share of South Asian residents
- We have a dataset of $m = 153$ practices and $n = 152$ areas and use the same priors as in the simulation study
- *Negative Binomial* likelihood, with an overdispersion parameter with prior: $\psi \sim \mathit{Gamma}(2, 0.2)$

Data Analysis



Probabilities $P(\rho_i > 1|\boldsymbol{y})$ that the relative risk in each exceeds 1 among posterior samples under the CAR-MM prior.

Conclusions and further work

- The CAR-MM allows two parameterisations, although only the *Post* is recommended
- The framework is general enough to encompass other applications beyond spatial misalignment (see Petrof et al., 2020) [8]
- The *Post* parameterised CAR-MM with informative priors is well calibrated

Further work:

- Possible Spatio-temporal extensions
- Explore more in depth the partial correlation structure of the *Inverse* parameterisation
- Scaling up the MCMC for larger m and n

Links

Bibliography - 1

1. Browne W.J., Goldstein H., Rasbash J., *Multiple membership multiple classification (MMMC) models*, Statistical Modelling, **1(2)**: 103-124 (2001).

2. Besag, J., *Spatial Interaction and the Statistical Analysis of Lattice Systems*, Journal of the Royal Statistical Society: Series B **36**: 192-225, 83 (1974).

3. Cole, D., *Parameter Redundancy and Identifiability*, Chapman & Hall/CRC Interdisciplinary Statistics (2020)

4. Talts S., Betancourt M., Simpson D., Vehtari A., Gelman A., *Validating Bayesian Inference Algorithms with Simulation-Based Calibration*, arXiv, 1804.06788 (2018)

5. Stan Development Team, *RStan: the R interface to Stan*, R package version 2.21.5 (2022)

6. Joseph M., *Exact sparse CAR models in Stan*, Stan Case Studies 3 (2016)

Bibliography - 2

7. Gramatica, M., Congdon, P. and Liverani, S., *Bayesian modelling for spatially misaligned health areal data: A multiple membership approach*, Journal of the Royal Statistical Society: Series C **70**: 645-666 (2021)

8. Petrof, O, Neyens, T, Nuyts, V, Nackaerts, K, Nemery, B, Faes, C. *On the impact of residential history in the spatial analysis of diseases with a long latency period: A study of mesothelioma in Belgium*, Statistics in Medicine **39**: 3840– 3866 (2020)