

# Exploratory analysis of accesses to support centers for gender-based violence in Apulia

## Data

The dataset employed regards the counts of accesses to gender-based violence support centers in the Apulia region by residence municipality of the women victims of violence in 2021-2024. R codes to generate the dataset are in the R script posted here which this report is based on.

Here, we only take into account the violence reports which support centers actually take charge of, at the risk of underestimating the counts of gender-based violence cases. This choice is driven by the need of avoiding duplicated records, since e.g. it may happen that a support center redirects a victim to another support center.

In order to avoid singletons, i.e. municipalities with no neighbours, in the spatial structure of the dataset, the Tremiti Islands need to be removed from the list of municipalities included (0 accesses recorded so far).

Therefore, the municipality-level dataset in scope consists of 256 areas.

We can only take into account the accesses to support centers for which the origin municipality of victims is reported; therefore the total count of accesses in scope is 1477, 1516, 1822 and 1778 for the reference years respectively:

Here, we plot the log-access rate per residence municipality, i.e. the logarithm of the ratio between access counts and female population. Blank areas correspond to municipalities from which zero women accessed support centers (82 municipalities).

## Covariates

Our target is explaining the number of accesses to support centers,  $y$ , defined at the municipality level, on the basis of a set of candidate known variables.  $y$  is modelled with simple Poisson regression.

We have at disposal a number of candidate explanatory variables, which include the distance of a municipality from the closest support center and a set of variables measuring social vulnerability under different dimensions; these latter covariates are provided by the ISTAT and are among the components of the Municipality Frailty Index (MFI). A more detailed description of MFI components is in this excel metadata file.

All covariates are quantitative variables and to ease model interpretation are scaled to have null mean and unit variance.

- TEP\_th, i.e. the distance of each municipality from the closest municipality hosting a support center. Distance is measured by road travel time in minutes (acronym TEP stays for Tempo Effettivo di Percorrenza, i.e. Actual Travel Time). Since to the best of our knowledge the list of active support centers changed between 2022 and 2023, we employ the list of centers active until 2022 for 2021-2022 data, and the list of centers active in 2023 for 2023 data.
- AES, the distance from the closest infrastructural pole, always measured in travel time. Infrastructural poles are defined as municipalities or clusters of neighbouring municipalities provided with a certain endowment in health, education and transport infrastructure (more details can be found, e.g., here).
- MFI, i.e. the decile of municipality frailty index.

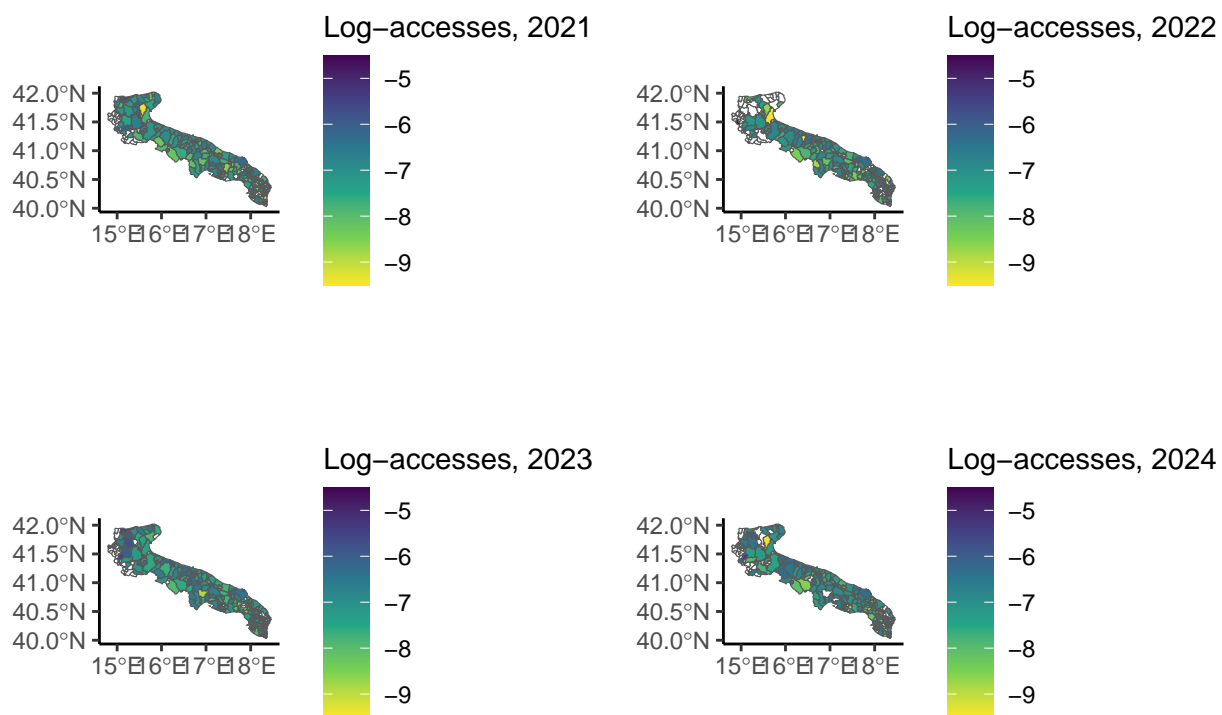


Figure 1: Log-access rate

- PDI, i.e. the dependency index, i.e. population either  $\leq 20$  or  $\geq 65$  years over population in  $[20 - 64]$  years.
- ELL, i.e. the proportion of people aged  $[25 - 54]$  with low education.
- ERR, i.e. employment rate among people aged  $[20 - 64]$ .
- PGR, i.e. population growth rate with respect to 2011.
- UIS, i.e. the ventile of the density of local units of industry and services (where density is defined as the ratio between the counts of industrial units and population).
- ELI, i.e. the ventile of employees in low productivity local units by sector for industry and services.

First, we visualise the correlations among these explanatory variables:

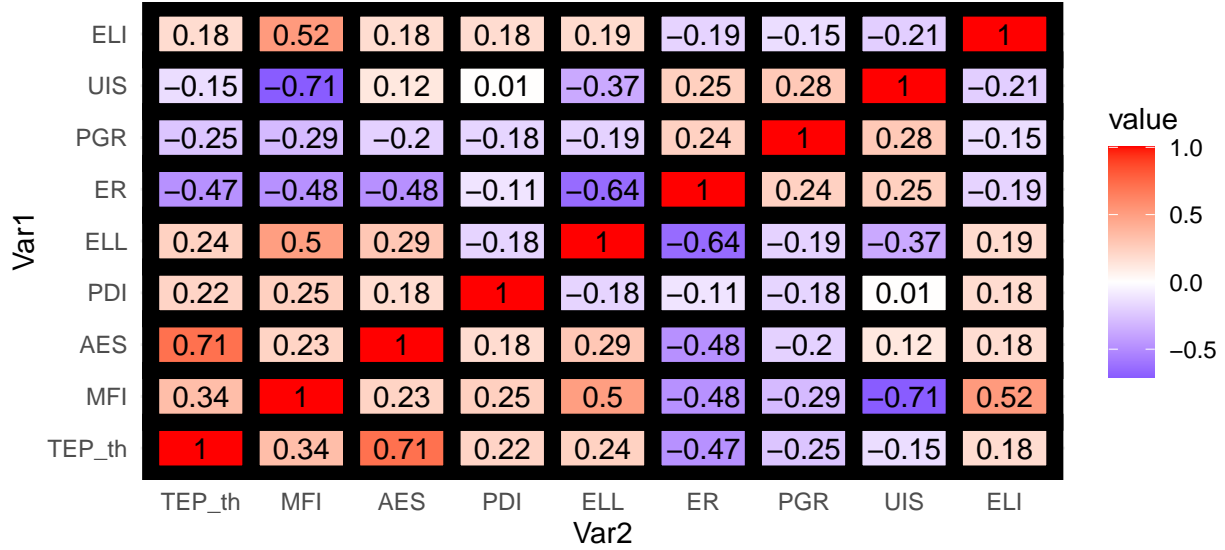


Figure 2: Correlations in explanatory variables

We see the correlation between the two distances is very high (0.72), and so is the correlation between the fragility index decile and the density of productive units.

In the first case, we drop the distance from the nearest infrastructural pole. In the latter we drop MFI, which is a combination of all covariates except for TEP\_th, and is a weakly informative choice.

## Nonspatial regression

We regress the counts of accesses  $y$  to support centers on the aforementioned explanatory variables. To interpret regression coefficients in an easier way, all covariates are scaled to zero mean and unit variance.

$$y_{it} \mid \eta_{it} \sim \text{Poisson}(E_{it} e^{\eta_{it}}) \quad \text{where} \quad \eta_{it} = X_{it}^{\top} \beta \quad (1)$$

Where  $X$  are the covariates defined earlier,  $\beta$  are covariate effects, and  $E_{it}$  is the female population aged  $\geq 15$  in municipality  $i$  and year  $t$ .

To gain more insight on the role of all explanatory variables we show the posterior summaries of the regression model; the posterior distribution is approximated with the INLA (Van Niekerk et al. 2023).

% latex table generated in R 4.5.0 by xtable 1.8-4 package % Mon Aug 18 09:26:40 2025

Effect	Mean	Sd	0.025quant	0.975quant
Int_2021	-7.333	0.030	-7.392	-7.274
Int_2022	-7.305	0.030	-7.364	-7.247
Int_2023	-7.128	0.028	-7.183	-7.073
Int_2024	-7.152	0.028	-7.208	-7.097
TEP_th	-0.227	0.017	-0.260	-0.195
ELI	-0.088	0.015	-0.117	-0.059
PGR	0.017	0.020	-0.021	0.056
UIS	0.040	0.016	0.009	0.071
ELL	-0.110	0.021	-0.150	-0.070
PDI	-0.065	0.021	-0.106	-0.025
ER	-0.212	0.022	-0.256	-0.169

- **TEP\_th**: The distance from the closest support center appears to play an important role. The easiest interpretation is that the physical distance represents a barrier to violence reporting. This is quite intuitive if we think of the material dynamics of reporting gender-based violence: one could reasonably expect violent men to prevent their partners to take a long rout and come out to report the violence suffered.
- **ELI**: The (ventile of the distribution of the) share of employees in low productivity economic units is a clear indicator of (relative) economic underdevelopment. The most naive interpretation would be that in underdeveloped areas reporting gender violence is somewhat harder than in developed ones.
- **PGR**: The association with population growth rate does not appear to be significantly different from zero.
- **UIS**: The (ventile of the distribution of the) density of production units appears to have a positive association with VAW reporting, though it is not very high in absolute value if compared to coariates such as TEP\_th.
- **ELL**: The association with the proportion of people with low educational level has negative sign and is high in absolute value. The interpretation seems quite easy: cultural development, in general, would encourage reporting violence.
- **PDI**: The association with population dependency index is negative, and its interpretation is ambiguous as well since the proportion of old or very young people can be read as a limiting factor both for VAW incidence and reporting.
- **ER**: The association with employment rate is very strong and bears negative sign for 2021 and 2023 data.

## Spatial regression

**Exploratory analysis of residuals** We plot the log-residuals  $\varepsilon$  of the nonspatial regression models, defined as  $\varepsilon := \ln y_{it} - \ln \hat{y}_{it}$  being  $\hat{y}_{it}$  the fitted value.

Residuals may exhibit spatial structure. To assess it, we employ the Moran and Geary tests. Since

Please notice that log-residuals only take finite values across the municipalities whose female citizens have reported at least one case of violence.

Additionally, this set of municipalities may include some singletons, which we remove to assess the value of the Moran and Geary statistics. Thus, for each year we have defined the indexes set **nonzero\_con** as the set of municipalities from which at least one case of gender-based violence has been reported, *and* which have at least one neighbouring municipality from which at least one case of gender-based violence was reported as well. For brevity, we only show the standardised  $I$  values, which under the null hypothesis should be distributed as  $N(0, 1)$ . The Geary's test is also included for completeness.

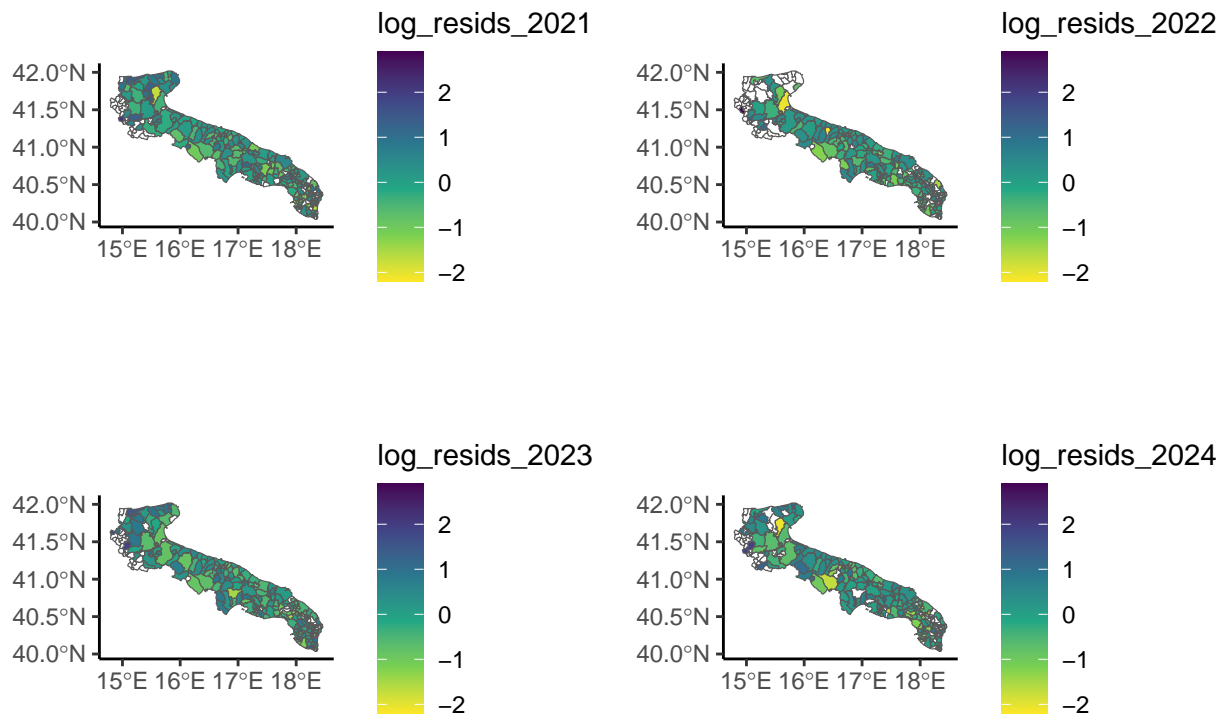


Figure 3: Log-residuals in GLM regression

Year	Test	Statistic_std	p.value
2021	Moran	1.78	0.04
2021	Geary	2.18	0.01
2022	Moran	2.20	0.01
2022	Geary	2.37	0.01
2023	Moran	4.85	0.00
2023	Geary	4.93	0.00
2024	Moran	3.94	0.00
2024	Geary	4.30	0.00

We find evidence for spatial autocorrelation. However, we must stress out that this result does not refer to all the regional territory, but only to a subset of all municipalities.

Based on the autocorrelation evidence, though it has only been assessed for a subset of all municipalities, we try implementing some simple spatial models by adding a conditionally autoregressive latent effect, say  $z$ , to the linear predictor

$$\eta_{it} = X_{it}^{\top} \beta + z_{it} \quad (2)$$

We test a total of four models, all of which have a prior distribution depending on the spatial structure of the underlying graph, in this case the Apulia region.

In the following, the area-specific latent field is denoted as  $z_i = (z_{i,2021} \ z_{i,2022} \ z_{i,2023} \ z_{i,2024})^{\top}$

We describe the spatial structure starting from municipalities neighbourhood, and introduce the neighbourhood matrix  $W$ , whose generic element  $w_{ij}$  takes value 1 if municipalities  $i$  and  $j$  are neighbours and 0 otherwise. For each  $i \in [1, n]$ ,  $d_i := \sum_{j=1}^n w_{ij}$  is the number of neighbours of  $i$ -th municipality. Please notice we have  $n = 256$ .

For all models, we define  $\Sigma$  as the covariance matrix of the latent effect  $z$ ; its off-diagonal entries describe dependence in  $z$  between different years. Spatial models are computed by approximating the marginal posteriors of interest via the Integrated Nested Laplace Approximation (INLA), adopting the novel Variational Bayes Approach (Van Niekerk et al. 2023). Before proceeding with model structure, we need a short digression on how R-INLA works. The INLA framework is hierarchical and is based on a distinction between latent Gaussian fields (GFs) and hyperparameters. In our case, the former include intercepts, covariate effects and the spatial latent field. In the INLA literature latent Gaussian fields are frequently labelled as  $x$  but we avoid doing so in order not to introduce notation ambiguity with respect to covariates. Hyperparameters in our case coincide with the parameters of  $z$ . Then, it is necessary to approximate the posterior distribution of both hyperparameters and latent GFs.

INLA follows a framework of numerical integration over a grid of hyperparameters, which is centered at the mode. Locating the posterior mode of hyperparameters is thus the first, crucial task to approximate all posterior distributions involved in the model. R-INLA employs an exploration strategy over the whole real axis (Rue, Martino, and Chopin 2009). For this reason, hyperparameters need be re-parametrised in such a way to be defined over all  $\mathbb{R}$ , and the priors are thus actually written on the so-called internal scale of hyperparameters. This approach may be hiding a trick if specific constraints are required on hyperparameters. For instance,  $\Sigma$  needs to be positive-definite and symmetric. If the hyperparameter exploration is carried out on the whole real axis, nothing ensures, in principle, that the values of the single entries in  $\Sigma$  ensure it is positive definite. For instance, consider the case the hyperparameter exploration procedure comes across this correlation matrix

$$\Sigma_{\text{ND}} = \begin{pmatrix} 1 & 0.4 & 0.9 \\ 0.4 & 1 & -0.4 \\ 0.9 & -0.4 & 1 \end{pmatrix}$$

It can be seen that  $\Sigma_{\text{ND}}$  has one negative eigenvalue, and is thus indefinite: it cannot be a valid correlation matrix and would cause trouble if, e.g., the covariance matrix is assigned a Wishart prior. A trivial solution

is to define the correlation matrix as a quadratic form, say  $\Sigma := BB^\top$ , and setting a prior distribution on  $B$ . Of course, it is convenient to define triangular  $B$  matrix - otherwise, both a high number of parameters would be required, and the prior on  $\Sigma$  could not be calculated (could it?) starting from the prior on  $B$  as the Jacobian matrix would *not* be squared.

Priors for spatial effects and hyperparameters are defined in this R script. The general structure follows the framework of the **bigDM** R package (Vicente et al. 2023).

**ICAR model** The Intrinsic CAR model is the simplest formulation among spatial autoregressive models. The conditional distribution of each value  $z_i \mid z_{-i}$  is:

$$z_i \mid z_{-i} \sim \mathcal{N} \left( \sum_{j=1}^n \frac{w_{ij}}{d_i} z_j, \frac{1}{d_i} \Lambda^{-1} \right) \quad (3)$$

And the joint prior distribution is:

$$z \mid \Sigma \sim \mathcal{N} (0, \Sigma \otimes (D - W)^+ ) \quad (4)$$

Where  $z$  is a  $256 \times 4$  matrix. Since the joint distribution of  $z$  is improper, a sum-to-zero constraint, i.e.  $\sum_{i=1}^n z_{it} = 0$  for  $t = 2021, 2022, 2023, 2024$  is required for identifiability.

**PCAR model** The intrinsic autoregressive model is relatively simple to interpret and to implement, while also requiring the minimum number of additional parameters (either the scale or the precision).

The drawback, however, is that we implicitly assume a deterministic spatial autocorrelation coefficient equal to 1. When the autocorrelation is weak, setting an ICAR prior may be a form of misspecification.

A generalisation of this model is the PCAR (proper CAR), which introduces an autocorrelation parameter  $\rho$ . We assign a penalised complexity (PC) prior to  $\rho$ , and, following the thumb-rule proposed by (Simpson et al. 2017) and (Riebler et al. 2016) for the Besag-York-Mollié model (see after) we assume *a priori* that  $\pi(\rho \leq \frac{1}{2}) = \frac{2}{3}$ .

$$z_i \mid z_{-i} \sim \mathcal{N} \left( \sum_{j=1}^n \rho \frac{w_{ij}}{d_i} z_j, \frac{1}{d_i} \Lambda^{-1} \right) \quad (5)$$

We show the posterior summary for the autocorrelation coefficient, obtained with the data at hand.

##	mean	sd	quant0.025	quant0.25	quant0.5	quant0.75	quant0.975
##	0.8496518	0.0698914	0.6808226	0.8107266	0.8615526	0.9011834	0.9500745

The credible interval for  $\rho$  is quite pushed towards unity, denoting the model estimates a strong spatial autocorrelation.

**Leroux model** As an alternative to take into account both structured and unstructured latent effects, we also test the Leroux autoregressive model (Leroux, Lei, and Breslow 2000); throughout this report, this model will be referred to as the LCAR. In this case, the local prior for  $z_i$  is

$$z_i \mid z_{-i} \sim \mathcal{N} \left( \sum_{j=1}^n \frac{\xi w_{ij}}{1 - \xi + \xi d_i} z_j, \Lambda^{-1} \frac{1}{1 - \xi + \xi d_i} \right) \quad (6)$$

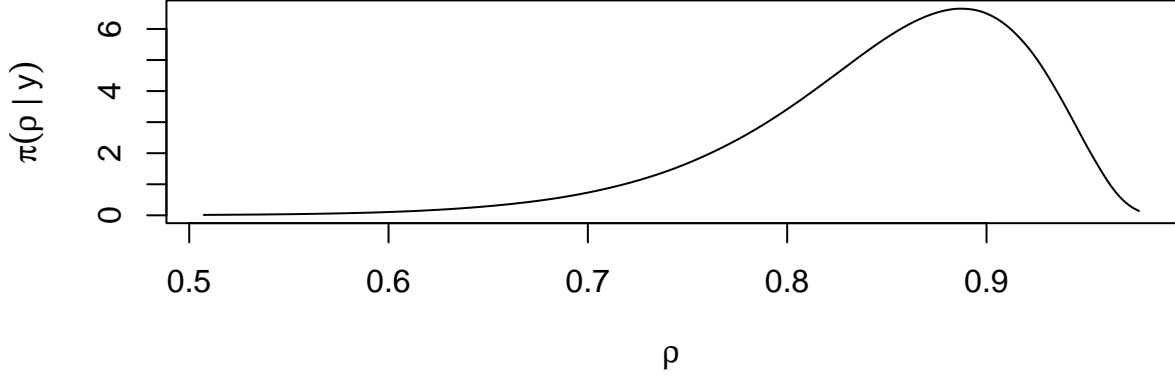


Figure 4: Posterior marginal of the PCAR autocorrelation parameter

Where  $\xi \in [0, 1]$  is labelled as the mixing parameter. A more interesting representation of the Leroux model is the joint prior

$$z \mid \Lambda, \xi \sim \mathcal{N}(0, [\Lambda \otimes (\xi L + (1 - \xi)I)]^{-1})$$

where  $L := D - W$  is the graph Laplacian matrix,  $W$  is the neighbourhood matrix and  $D$  is the corresponding degree matrix. We can clearly see how the mixing parameter allocates variability between two precision components, i.e. the Laplacian matrix for the spatial part and the identity matrix for the noise.

The drawback of this model is the scarce interpretability with respect to more sophisticated ones like the BYM.

**BYM model** Another popular model to control for both spatial autocorrelation and random noise is the Besag, York and Mollié model. We employ a simplified formulation, with a unique mixing parameter for all three years. Under this model the latent effect is defined as:

$$z = \sqrt{\phi}uM + \sqrt{1 - \phi}vM \quad (7)$$

Where:

- $u \sim \mathcal{N}(0, I_p \otimes L_{\text{scaled}})$  is an independent multivariate ICAR process whose precision matrix is scaled in order that the geometric mean of marginal variances (i.e. the diagonal entries of  $L^+$ ) is one
- $v \sim \mathcal{N}(0, I_p \otimes I_n)$  is a Standard Normal variable
- $M$  is a positive definite matrix such that  $M'M = \Lambda^{-1}$ . It is not necessarily the Cholesky factor of the scale parameter; in fact, a convenient but not unique way to define it may be  $M = D^{-\frac{1}{2}}E^\top$  where  $E$  and  $D$  are the eigenvector and eigenvalues matrices of  $\Lambda$  (Urdangarin et al. 2024).
- $\phi \in [0, 1]$  is the mixing parameter.

Please notice that in the case of our data,  $p = 3$  and  $n = 256$ . We assign a Uniform prior on  $\phi$  (but the PC-prior would be a more rigorous choice) and the usual Wishart prior to  $\Lambda$ .

The BYM model features a variance mixing parameter, whose interpretation is eased by allowing to scale the ICAR and IID components.



We briefly compare the three models in scope through the WAIC (Gelman, Hwang, and Vehtari 2014):

% latex table generated in R 4.5.0 by xtable 1.8-4 package % Mon Aug 18 09:31:41 2025

Model	WAIC	Eff_params
Null	4798.741	44.505
ICAR	3899.529	257.647
PCAR	3918.801	265.977
Leroux	3907.826	265.791
BYM	3880.151	257.395

As we can see, adding a spatial model is an improving element, not a waste of complexity.

Here we show some posterior summaries for  $\beta$  under the BYM model.

% latex table generated in R 4.5.0 by xtable 1.8-4 package % Mon Aug 18 09:31:42 2025

Mean	Sd	Q_0.025	Q_0.975
-7.439	0.049	-7.537	-7.344
-7.613	0.061	-7.737	-7.496
-7.199	0.048	-7.295	-7.105
-7.339	0.054	-7.447	-7.234
-0.204	0.046	-0.295	-0.113
-0.001	0.039	-0.078	0.077
0.077	0.043	-0.008	0.163
0.086	0.043	0.002	0.171
-0.228	0.052	-0.330	-0.126
-0.036	0.047	-0.128	0.056
-0.274	0.061	-0.393	-0.156

Estimations of  $\beta$  differ slightly from the nonspatial model and credibility intervals are generally wider due to increased uncertainty.

- **TEP\_th** The effect of the distance from the closest support center remains similar in mean and the interpretation is not altered.
- **ELI**: The effect of the incidence of low-productivity economic units is utterly negligible
- **PGR**: The association with population growth rate is positive as in the nonspatial model, though it is lesser in absolute value
- **UIS**: The association with the density of productive units is positive and double, on average, with respect to the nonspatial model. Interpreting this association is complex, as it bears opposite sign with the employment rate, which should be another indicator of economic development.
- **ELL**: The association with the incidence of low education levels, is even higher in mean than under the nonspatial model. We interpret this result as a strong *potential* impact of education on the chance that gender violence is reported.
- **PDI**: The effect of structural dependency index is utterly negligible, as for the nonspatial model.
- **ER**: The effect associated with employment rate is negative and relatively high in absolute value. How to interpret this finding? Employment rate is clearly an indicator of economic development, hence the easiest interpretation is that in more developed areas there is a higher chance that gender violence is reported.

We show the expected latent effects:

We additionally show the fitted values of the counts of accesses, rounded to the closest integer.

We show the posterior median of the marginal variances of  $z$ , i.e. the diagonal entries of  $\Sigma$ :

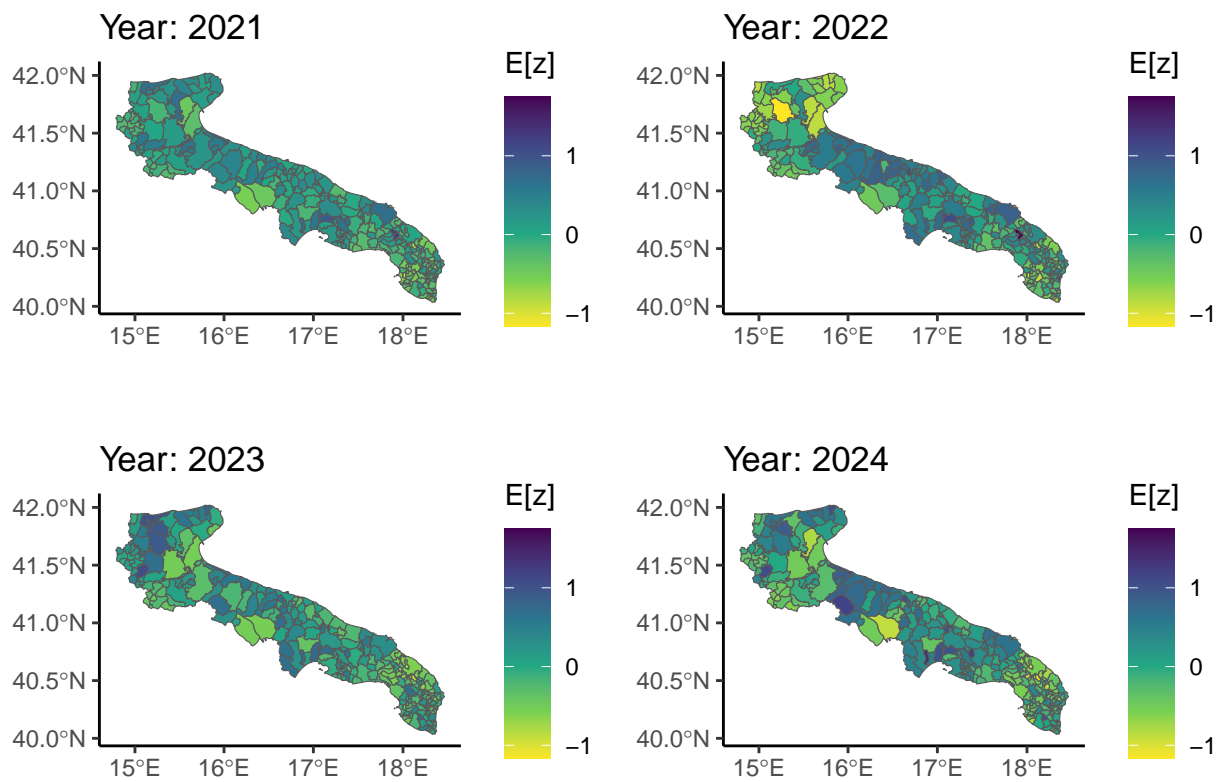


Figure 5: Estimated BYM latent effect

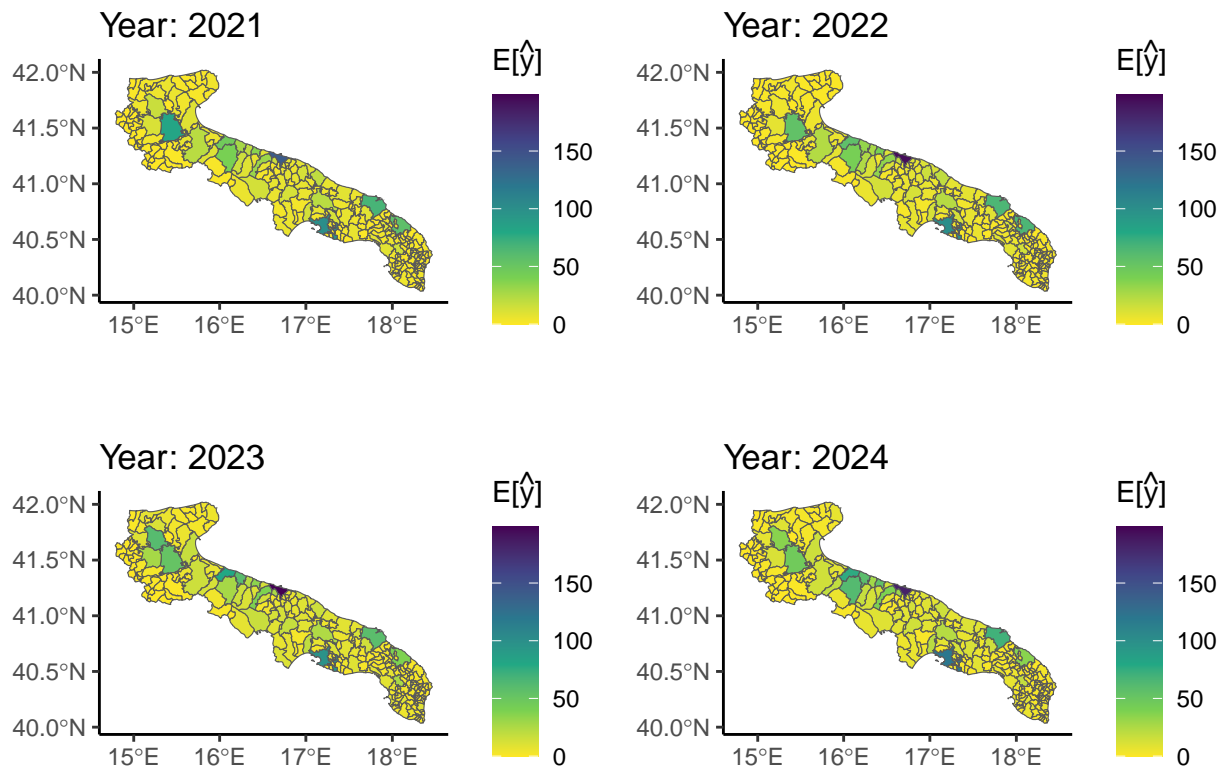


Figure 6: Fitted access counts, BYM model

```
## % latex table generated in R 4.5.0 by xtable 1.8-4 package
## % Mon Aug 18 09:31:53 2025
## \begin{table}[ht]
## \centering
## \begin{tabular}{lrrrrr}
## \hline
## Year & Mean & Sd & Q0.025 & Median & Q0.975 \\\
## \hline
## 2021 & 0.308 & 0.096 & 0.162 & 0.294 & 0.533 \\\
## 2022 & 0.541 & 0.142 & 0.300 & 0.528 & 0.853 \\\
## 2023 & 0.400 & 0.104 & 0.232 & 0.389 & 0.641 \\\
## 2024 & 0.562 & 0.135 & 0.330 & 0.550 & 0.858 \\\
## \hline
## \end{tabular}
## \end{table}
```

For more insight on the dependence structure between the three years, posterior summaries for correlations are shown in the following. In terms of point estimates, the only high correlation appears that of 2021-2022.

% latex table generated in R 4.5.0 by xtable 1.8-4 package % Mon Aug 18 09:31:54 2025

Years	Mean	Sd	Q0.025	Median	Q0.975
2021-22	0.763	0.087	0.566	0.775	0.899
2021-23	0.347	0.134	0.067	0.353	0.587
2021-24	0.482	0.127	0.212	0.492	0.706
2022-23	0.226	0.130	-0.045	0.232	0.468
2022-24	0.261	0.129	-0.006	0.264	0.502
2023-24	0.703	0.086	0.511	0.712	0.848

Lastly, we have a look at the mixing parameter under the Leroux and BYM models.

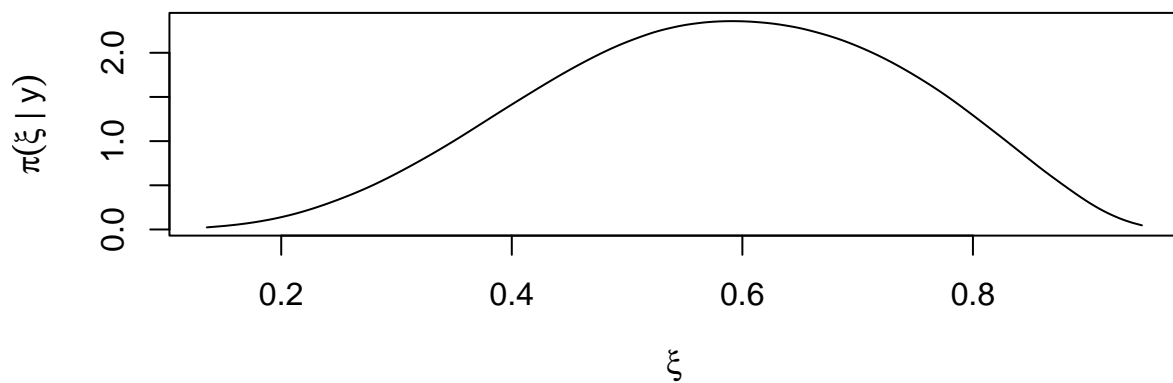


Figure 7: Posterior marginal of the LCAR mixing parameter

```
##          mean          sd quant0.025 quant0.25 quant0.5 quant0.75 quant0.975
## 0.5819261 0.1563537 0.2739076 0.4692732 0.5859589 0.6998273 0.8615089
```

As we see, the Leroux mixing parameter is not particularly high, but interpreting it properly is hampered by the difficulty in scaling the precision matrix, which is a drawback of non-intrinsic models.

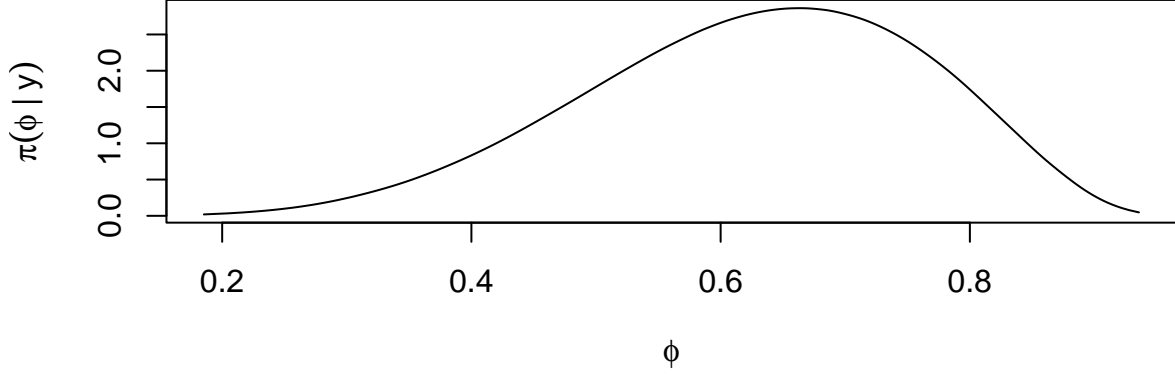


Figure 8: Posterior marginal of the BYM mixing parameter

##	mean	sd	quant0.025	quant0.25	quant0.5	quant0.75	quant0.975
##	0.6270466	0.1358362	0.3422842	0.5345197	0.6364752	0.7288332	0.8593818

#### M-models Until now, we have been assuming the hyperparameter controlling for the strength spatial association is constant across the three years. We try to relax this hypothesis by allowing for year-specific hyperparameters, which can be achieved by using a more flexible class of multivariate models, i.e. the M-models (Botella-Rocamora, Martinez-Beneito, and Banerjee 2015).

Consider the scale parameter  $\Sigma$ , i.e. the marginal variance of the latent effect  $z$ , and let us define the definite-positive matrix  $M$ , such that  $M^\top M = \Sigma$ .

Starting with the PCAR model, if we allow the autocorrelation parameter  $\rho$  to change over time, we introduce the matrix-valued parameter  $R := \text{diag}(\rho_{2021}, \rho_{2022}, \rho_{2023})$  and the model becomes thus:

$$z \mid \Sigma, R \sim \mathcal{N}(0, (M^\top \otimes I_n)[(I_p \otimes D) - (R \otimes W)]^{-1}(M \otimes I_n)) \quad (8)$$

Where  $n = 256$  and  $p = 4$ , and the remainder of notation is equal to previous sections. Since the matrix  $M$  has dimension  $p^2$  and is not symmetric, modelling it directly would become burdensome. We therefore opt to make directly inference on  $\Sigma$ . In line with (Vicente et al. 2023), we employ the Bartlett factorisation to  $\Sigma$ , which briefly means that we set  $\Sigma := BB^\top$ , where  $B$  is a lower-triangular matrix whose off-diagonal elements are assigned a Normal prior and the squares of its diagonal are assigned a  $\chi^2$  distribution. This is equivalent to the assumption that  $\Sigma$  follows a Wishart prior; for a proof see e.g. (Kabe 1964).

As we previously mentioned, we loosely follow the workflow underlying the **bigDM** R package (Vicente et al. 2023), which implements the M-model formulation of the PCAR, LCAR and BYM models.

The autoregressive and mixing parameters follow a multivariate and independent PC-prior; for the time being, we keep assuming  $\pi(\psi \leq \frac{1}{2}) = \frac{2}{3}$ , where  $\psi$  is a generic hyperparameter. More detail on PC-prior definition can be found in my PhD thesis or in the notes of the shared overleaf project.

Likewise, we test the LCAR model, which under the M-model formulation features a matrix-valued mixing parameter  $\Xi := \text{diag}(\xi_{2021}, \xi_{2022}, \xi_{2023})$

$$z \mid \Sigma, R \sim \mathcal{N}(0, (M^\top \otimes I_n)[\Xi \otimes (D - W) + (I_p - \Xi) \otimes I_n]^{-1}(M \otimes I_n)) \quad (9)$$

Lastly, we propose the M-model extension of the BYM.

If we define the matrix-valued mixing parameter  $\Phi := \text{diag}(\phi_1, \phi_2, \dots, \phi_p)$  and  $\bar{\Phi} := \text{diag}(1 - \phi_1, 1 - \phi_2, \dots, 1 - \phi_p) = (I_p - \Phi)$ , with  $p = 3$ , the BYM field  $\mathbf{Y}$  is given by:

$$Z = U\Phi^{\frac{1}{2}}M + V\bar{\Phi}^{\frac{1}{2}}M \quad (10)$$

Where

- $U = (U_1, U_2, \dots, U_p)$  is a multivariate and independent ICAR field, such that  $U_j \sim \mathcal{N}_n(0, L^+)$  for  $j = 1, 2, \dots, p$ , and  $L^+$  is the pseudoinverse of the scaled graph Laplacian matrix
- $V = (V_1, V_2, \dots, V_p)$  is a collection of *iid* random Normal vectors, such that  $V_j \sim \mathcal{N}_n(0, I_n)$  for  $j = 1, 2, \dots, p$ .
- $M$  is a generic  $p \times p$  full-rank matrix such that  $M^\top M = \Sigma$ , being  $\Sigma$  is the scale parameter. For instance,  $M$  can be defined as  $D^{1/2}E^\top$ , where  $D$  is the diagonal matrix of the eigenvalues of  $\Sigma$  and  $E$  is the corresponding eigenvectors matrix.

```
cav_MMBYM_inla <- inla(
  N_ACC ~ 0 + Y_2021 + Y_2022 + Y_2023 + Y_2024 +
  TEP_th + ELI + PGR + UIS + ELL + PDI + ER+
  f(ID, model = inla.MMBYM.model(k = 4, W = W_con, PC= T, df = 6) ,
    extraconstr = constr.BYM),
  offset = log(nn), #control.inla = list(h = 1e-5),
  family = "poisson", data = dd,
  num.threads = 1, control.compute = list(internal.opt = F, cpo = T, waic = T, config = T),
  verbose = F)
```

Model comparison follows: % latex table generated in R 4.5.0 by xtable 1.8-4 package % Mon Aug 18 10:07:26 2025

Model	WAIC	Eff_params
Null	4798.741	44.505
ICAR	3899.529	257.647
PCAR	3897.151	263.679
Leroux	3877.668	258.443
BYM	3871.098	252.077

Here we show some posterior summaries for  $\alpha$  under the BYM model.

% latex table generated in R 4.5.0 by xtable 1.8-4 package % Mon Aug 18 10:07:26 2025

We then show the expected latent effects.

We additionally show the fitted values of the counts of accesses, rounded to the closest integer.

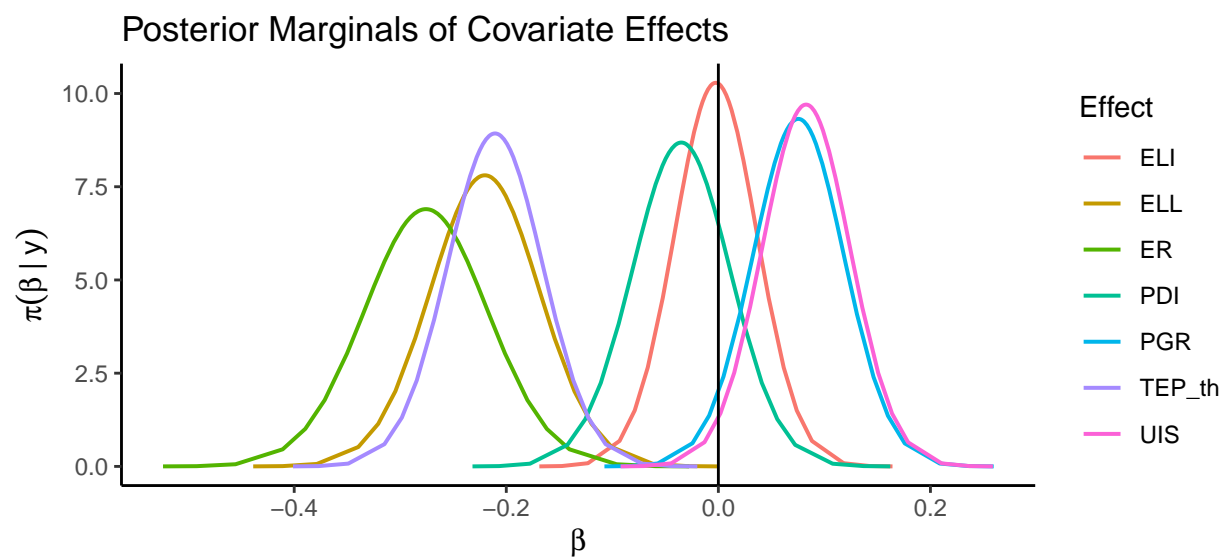


Figure 9: Posterior marginals of covariate effects, M-model BYM

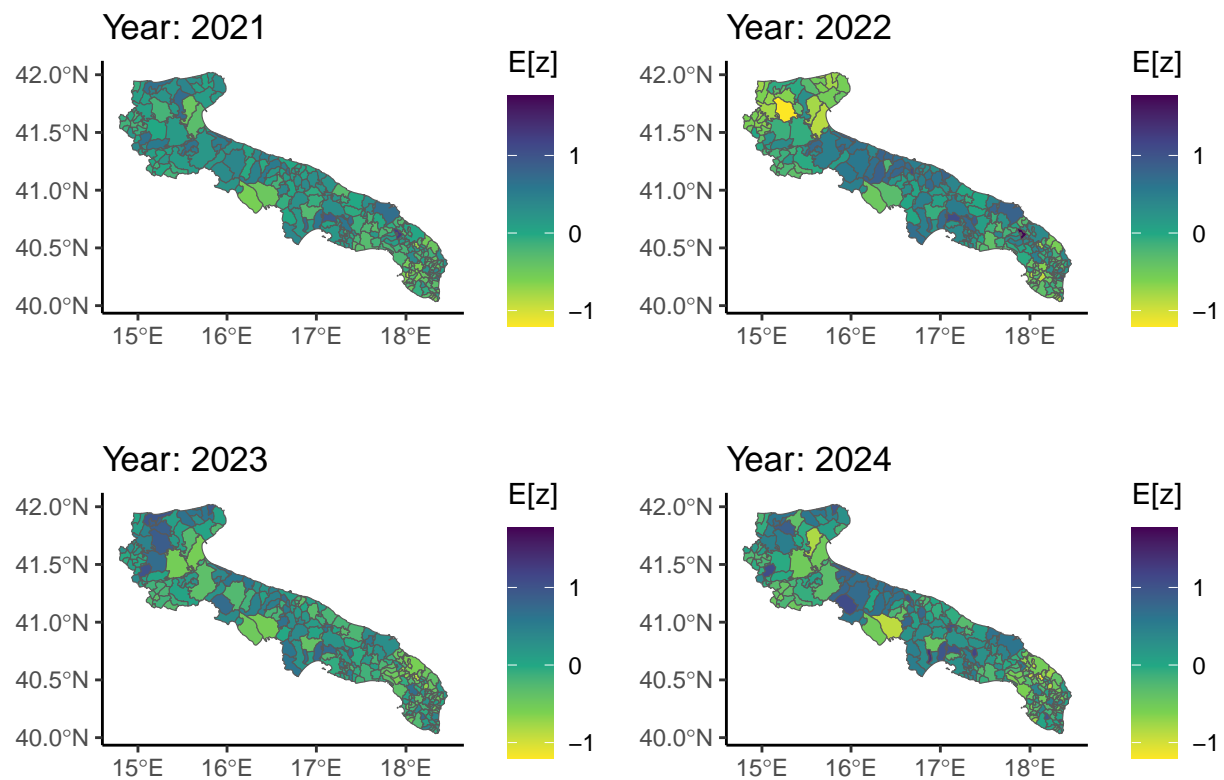
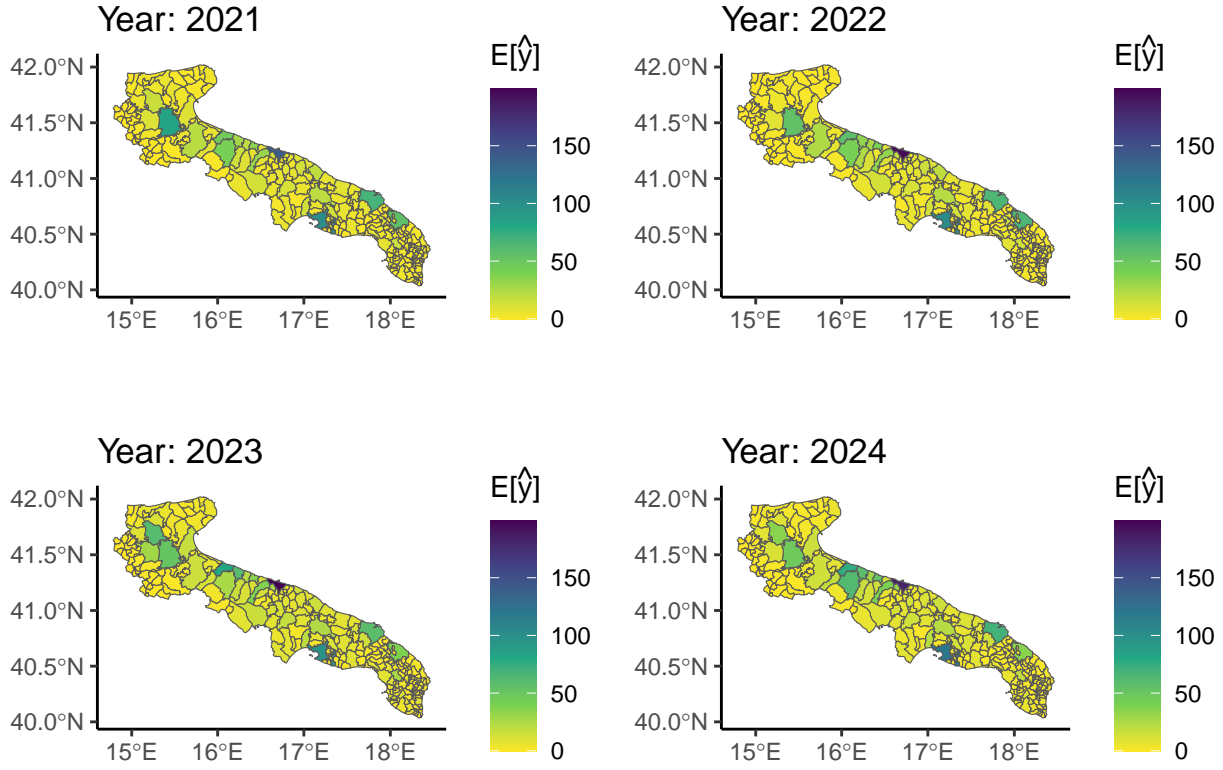


Figure 10: Estimated BYM latent effect using the M-model formulation

Effect	Mean	Sd	Q_0.025	Q_0.975
Y_2021	-7.434	0.047	-7.527	-7.341
Y_2022	-7.609	0.058	-7.724	-7.496
Y_2023	-7.194	0.047	-7.286	-7.103
Y_2024	-7.328	0.053	-7.434	-7.225
TEP_th	-0.211	0.045	-0.298	-0.123
ELI	-0.002	0.039	-0.079	0.074
PGR	0.076	0.043	-0.009	0.160
UIS	0.083	0.041	0.002	0.164
ELL	-0.220	0.051	-0.321	-0.120
PDI	-0.035	0.046	-0.125	0.056
ER	-0.276	0.058	-0.390	-0.162



Differences with respect to the scalar-mixing model are hardly noticeable in terms of both  $\beta$ ,  $\eta$  and  $z$ . Lastly, we have a look at the autoregressive (PCAR) and mixing (LCAR, BYM) parameters.

% latex table generated in R 4.5.0 by xtable 1.8-4 package % Mon Aug 18 10:07:36 2025

Here we plot the marginal posterior of the BYM variance mixing parameter.

The interpretation is that while spatial association is moderate in 2021 data, the spatial distribution of 2023 data is dominated by noise; 2022 and 2024 data instead display a rather strong spatial association. For 2023 data, the PC prior penalises indeed the departure from the nonspatial model quite strongly.

**Appendix: sparse parametrisation of the BYM** Recall the BYM as defined in equation 10; the prior variance of  $Z$  is:

$$\text{Var}[\text{vec}(Z) \mid \Sigma, \Phi] = (M^\top \otimes I_n) \text{diag}(S_1, \dots, S_p) (M \otimes I_n)$$



Model	Year	Mean	Sd	Q0.50	Q0.025	Q0.975
PCAR	2021	0.449	0.223	0.070	0.445	0.858
PCAR	2022	0.940	0.056	0.783	0.958	0.992
PCAR	2023	0.653	0.173	0.272	0.676	0.916
PCAR	2024	0.710	0.187	0.270	0.750	0.957
LCAR	2021	0.398	0.113	0.202	0.389	0.637
LCAR	2022	0.549	0.190	0.182	0.556	0.879
LCAR	2023	0.293	0.150	0.074	0.267	0.639
LCAR	2024	0.972	0.008	0.953	0.974	0.986
BYM	2021	0.574	0.164	0.264	0.573	0.873
BYM	2022	0.695	0.177	0.273	0.735	0.931
BYM	2023	0.111	0.118	0.005	0.070	0.449
BYM	2024	0.670	0.182	0.266	0.697	0.938

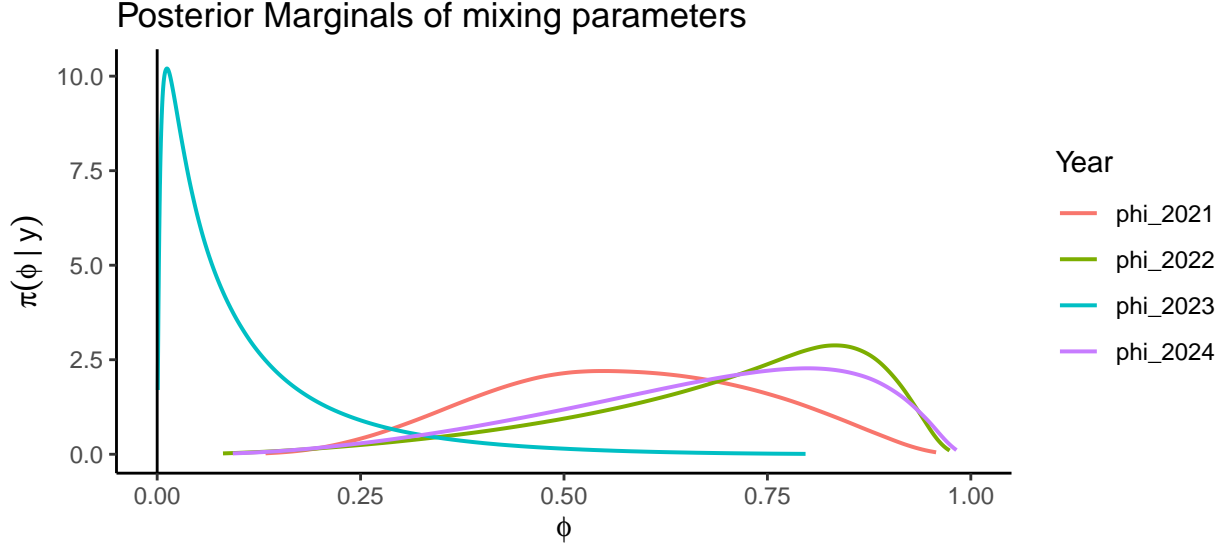


Figure 11: Posterior marginal of the BYM mixing parameters

Where each diagonal block  $S_j$  is the variance of  $U_j\Phi + V_j\bar{\Phi}$ , i.e.  $S_j = \phi_j L^+ + (1 - \phi_j)I_n$ .

Extending the approach of (Riebler et al. 2016) to the multivariate case, we know that:

$$\mathbb{E}[\text{vec}(Z) \mid U, \Phi, \Sigma] = \text{vec}(U\Phi^{\frac{1}{2}}M) = [(M^\top \Phi^{\frac{1}{2}}) \otimes I_n] \text{vec}(U)$$

and similarly

$$\text{Var}[\text{vec}(Z) \mid U, \Phi, \Sigma] = [(M^\top \bar{\Phi}^{\frac{1}{2}}) \otimes I_n] \mathbb{E}[\text{vec}(V)\text{vec}(V)^\top] [(\bar{\Phi}^{\frac{1}{2}}M) \otimes I_n] = (M^\top \bar{\Phi}M) \otimes I_n$$

The distribution of  $Z \mid U, \Sigma, \Phi$  then reads:

$$\begin{aligned}
& -2 \ln \pi (\text{vec}(Z) \mid U, \Sigma, \phi) \text{vec}(Z)^\top \left[ \left( M^{-1} \bar{\Phi}^{-1} M^{-1\top} \right) \otimes I_n \right] \text{vec}(Z) \\
& - 2 \text{vec}(Z)^\top \left[ \left( M^{-1} \bar{\Phi}^{-1} M^{-1\top} \right) \otimes I_n \right] \left[ \left( M^\top \Phi^{\frac{1}{2}} \right) \otimes I_n \right] \text{vec}(U) \\
& + \text{vec}(U)^\top \left[ \left( \Phi^{\frac{1}{2}} M \right) \otimes I_n \right] \left[ \left( M^{-1} \bar{\Phi}^{-1} M^{-1\top} \right) \otimes I_n \right] \left[ \left( M^\top \Phi^{\frac{1}{2}} \right) \otimes I_n \right] \text{vec}(U) \\
& = C + \text{vec}(Z)^\top \left[ \left( M^{-1} \bar{\Phi}^{-1} M^{-1\top} \right) \otimes I_n \right] \text{vec}(Z) \\
& - 2 \text{vec}(Z)^\top \left[ \left( M^{-1} \bar{\Phi}^{-1} \Phi^{\frac{1}{2}} M^\top \right) \otimes I_n \right] \text{vec}(U) \\
& + \text{vec}(U)^\top \left[ \left( \Phi \bar{\Phi}^{-1} \right) \otimes I_n \right] \text{vec}(U)
\end{aligned}$$

Now, for brevity let us define the following  $p \times p$  matrices:

$$q_{11} := M^{-1} \bar{\Phi}^{-1} M^{-1\top}; \quad q_{12} := M^{-1} \bar{\Phi}^{-1} \Phi^{\frac{1}{2}}; \quad q_{22} := \Phi \bar{\Phi}^{-1}$$

Hence

$$\begin{aligned}
-2 \ln \pi (\text{vec}(Z) \mid U, \Sigma, \Phi) &= C + \text{vec}(Z)^\top (q_{11} \otimes I_n) \text{vec}(Z) \\
& - 2 \text{vec}(Z)^\top (q_{12} \otimes I_n) \text{vec}(U) \\
& + \text{vec}(U)^\top (q_{22} \otimes I_n) \text{vec}(U)
\end{aligned}$$

Then, we have

$$\begin{aligned}
-2 \ln \pi (\text{vec}(Z), \text{vec}(U) \mid \Sigma, \phi) &= C + \text{vec}(Z)^\top (q_{11} \otimes I_n) \text{vec}(Z) \\
& - 2 \text{vec}(Z)^\top (q_{12} \otimes I_n) \text{vec}(U) \\
& + \text{vec}(U)^\top (q_{12} \otimes I_n + I_p \otimes L) \text{vec}(U)
\end{aligned}$$

Hence, with some straightforward algebra, it can be concluded that:

$$\begin{pmatrix} \text{vec}(Z) \\ \text{vec}(U) \end{pmatrix} \sim \mathcal{N}_{2np} \left( 0, \begin{pmatrix} q_{11} \otimes I_n & -q_{12} \otimes I_n \\ -q_{12}^\top \otimes I_n & q_{22} \otimes I_n + I_p \otimes L \end{pmatrix}^{-1} \right) \quad (11)$$

Which generalises to the multivariate case the sparse precision derived by (Riebler et al. 2016).

Botella-Rocamora, Paloma, Miguel A Martinez-Beneito, and Sudipto Banerjee. 2015. “A Unifying Modeling Framework for Highly Multivariate Disease Mapping.” *Statistics in Medicine* 34 (9): 1548–59.

Gelman, Andrew, Jessica Hwang, and Aki Vehtari. 2014. “Understanding Predictive Information Criteria for Bayesian Models.” *Statistics and Computing* 24 (6): 997–1016. <https://doi.org/10.1007/S11222-013-9416-2>.

Kabe, Dattatraya G. 1964. “A Note on the Bartlett Decomposition of a Wishart Matrix.” *Journal of the Royal Statistical Society Series B: Statistical Methodology* 26 (2): 270–73. <https://doi.org/https://doi.org/10.1111/j.2517-6161.1964.tb00558.x>.

Leroux, Brian G., Xingye Lei, and Norman Breslow. 2000. “Estimation of Disease Rates in Small Areas: A New Mixed Model for Spatial Dependence.” In *Statistical Models in Epidemiology, the Environment, and Clinical Trials*, edited by M. Elizabeth Halloran and Donald Berry, 179–91. New York, NY: Springer New York. [https://doi.org/https://doi.org/10.1007/978-1-4612-1284-3\\_4](https://doi.org/https://doi.org/10.1007/978-1-4612-1284-3_4).

Riebler, Andrea, Sigrunn H Sørbye, Daniel Simpson, and Håvard Rue. 2016. “An Intuitive Bayesian Spatial Model for Disease Mapping That Accounts for Scaling.” *Statistical Methods in Medical Research* 25 (4): 1145–65.

Rue, Haavard, Sara Martino, and Nicholas Chopin. 2009. “Approximate Bayesian Inference for Latent Gaussian Models by Using Integrated Nested Laplace Approximations.” *Journal of the Royal Statistical Society Series B: (Methodological)* 71 (2): 319–92. <https://doi.org/10.1111/j.1467-9868.2008.00700.x>.

- Simpson, Daniel, Haavard Rue, Andrea Riebler, Thiago G. Martins, and Sigrunn H. Sørbye. 2017. “Penalising Model Component Complexity: A Principled, Practical Approach to Constructing Priors.” *Statistical Science* 32 (1): 1–28. <https://doi.org/10.1214/16-STS576>.
- Urdangarin, Arantxa, Tomas Goicoa, T. Kneib, and M. D. Ugarte. 2024. “A Simplified Spatial+ Approach to Mitigate Spatial Confounding in Multivariate Spatial Areal Models.” *Spatial Statistics* 59: 100804. <https://doi.org/10.1016/j.spasta.2023.100804>.
- Van Niekerk, Janet, Elias Krainski, Denis Rustand, and Haavard Rue. 2023. “A New Avenue for Bayesian Inference with INLA.” *Computational Statistics and Data Analysis* 181. <https://doi.org/10.1016/j.csda.2023.107692>.
- Vicente, Gonzalo, Aritz Adin, Tomás Goicoa, and María Dolores Ugarte. 2023. “High-Dimensional Order-Free Multivariate Spatial Disease Mapping.” *Statistics and Computing* 33 (5): 104.