

# Bayesian inference with INLA - 2.0

Janet van Niekerk

Janet.vanNiekerk@kaust.edu.sa

King Abdullah University of  
Science and Technology



August 2024



# QR code for the slides



[https://github.com/JanetVN1201/SA\\_US\\_24](https://github.com/JanetVN1201/SA_US_24)



# Outline

- 1 Introduction to INLA
- 2 Posterior inference with INLA
  - INLA 1.0
- 3 Low-rank VB correction
- 4 INLA 2.0
- 5 Examples
  - Illustrative examples
  - (Non-stationary) disease mapping
  - Geostatistics
  - Barrier model - non-stationary Matern field
  - Dementia study - SPDE on 3D
- 6 Discussion



# BayesComp group at KAUST





# Bayesian inference

Data  $\mathbf{y}$  (with covariates  $\mathbf{Z}$ ), depend on  $\mathbf{X}$  and  $\boldsymbol{\theta}$  such that,  $E[Y] = h(\mathbf{A}(\mathbf{Z})\mathbf{X})$ .

Bayes' theorem:

$$\begin{aligned} q(\mathbf{X}, \boldsymbol{\theta} | \mathbf{y}) &\propto L(\mathbf{y} | \mathbf{X}, \boldsymbol{\theta}) p(\mathbf{X}, \boldsymbol{\theta}) \\ \text{Posterior} &\propto \text{Likelihood} \times \text{Prior} \end{aligned}$$



# Computational aspects

- Analytical methods - conjugacy (pre-computer era)
- Approximate methods - Laplace (can be inaccurate)
- Exact methods - MCMC (very slow for complex models or large data)

Now, due to computing resources approximate methods are gaining popularity - INLA, VB, EP etc.

INLA - 2009 [Rue et al., 2009]

2021+ [Van Niekerk et al., 2023]

HPC [Gaedke-Merzhäuser et al., 2023]



# Model definition - GAMM

Suppose we have response data  $\mathbf{y}_{n \times 1}$  (conditionally independent) with density function  $\pi(y|\mathbf{X}, \boldsymbol{\theta})$  and link function  $h(\cdot)$ , that is linked to some covariates  $\mathbf{Z}$  through linear predictors

$$\boldsymbol{\eta}_n = \beta_0 + \mathbf{Z}_\beta \boldsymbol{\beta} + \sum f^k(\mathbf{Z}_f) = \mathbf{A}\mathbf{X}$$

The inferential aim is to estimate the latent field  $\mathbf{X}_m = \{\beta_0, \boldsymbol{\beta}, \mathbf{f}\}$ , and  $\boldsymbol{\theta}$ .



# GAMM → LGM

Assume

$$\boldsymbol{X}|\boldsymbol{\theta} \sim N(\boldsymbol{0}, \boldsymbol{Q}(\boldsymbol{\theta})^{-1})$$

where  $\boldsymbol{Q}(\boldsymbol{\theta})$  is a sparse matrix ( $\boldsymbol{X}$  is a GMRF).

$p(\boldsymbol{X}, \boldsymbol{\theta}) = p(\boldsymbol{X}|\boldsymbol{\theta})p(\boldsymbol{\theta})$  and  $p(\boldsymbol{\theta})$  can be non-Gaussian.



# Why is INLA so accurate and so fast?

- LGM structure
- Sparse precision matrix
- Specialized matrix algebra for sparse matrices
- NEW: VB (low-rank) correction [van Niekerk and Rue, 2024]

Use precision matrix instead of covariance matrix → natural occurrence



# How common are sparse $Q(\theta)$ ?

Consider an AR(1) model..



# AR(1) example



# AR(1) example



# Ingredients

Likelihood -  $\prod_{i=1}^n \pi(y_i | \mathbf{X}, \boldsymbol{\theta})$

Prior for the latent -  $\pi(\mathbf{X} | \boldsymbol{\theta})$

Prior for the hyperparameters -  $\pi(\boldsymbol{\theta})$

Goal:

- $q(X_j | \mathbf{y})$
- $q(\theta_k | \mathbf{y})$



# Posterior approximations by INLA

$$\begin{aligned}\pi(\boldsymbol{X}, \boldsymbol{\theta}, \boldsymbol{y}) &= \pi(\boldsymbol{\theta})\pi(\boldsymbol{X}|\boldsymbol{\theta}) \prod_{i=1}^n \pi(y_i | (\boldsymbol{AX})_i, \boldsymbol{\theta}) \\ \tilde{q}(\boldsymbol{\theta}|\boldsymbol{y}) &\propto \frac{\pi(\boldsymbol{X}, \boldsymbol{\theta}, \boldsymbol{y})}{\pi_G(\boldsymbol{X}|\boldsymbol{\theta}, \boldsymbol{y})} \Big|_{\boldsymbol{X}=\mu(\boldsymbol{\theta})} \\ \tilde{q}(\theta_j|\boldsymbol{y}) &= \int \tilde{\pi}(\boldsymbol{\theta}|\boldsymbol{y}) d\boldsymbol{\theta}_{-j} \\ \tilde{q}(\boldsymbol{X}_j|\boldsymbol{y}) &= \int \tilde{q}(\boldsymbol{X}_j|\boldsymbol{\theta}, \boldsymbol{y}) \tilde{q}(\boldsymbol{\theta}|\boldsymbol{y}) d\boldsymbol{\theta},\end{aligned}$$

$\tilde{q}(\boldsymbol{X}_j|\boldsymbol{\theta}, \boldsymbol{y})$  depends on the approximation used, for Gaussian it is straightforward for the Laplace approximation we do another Gaussian approximation to  $\tilde{q}(\boldsymbol{X}_{-j}|\boldsymbol{\theta}, \boldsymbol{y})$ .



# INLA 1.0

- $\boldsymbol{X} = \{\boldsymbol{\eta}, \beta_0, \boldsymbol{\beta}, \boldsymbol{f}\}$
- For each  $j$ ,

$$\tilde{q}(x_j | \boldsymbol{\theta}, \mathbf{y}) = \frac{\pi(\boldsymbol{X}, \boldsymbol{\theta}, \mathbf{y})}{\pi_G(\boldsymbol{X}_{-j} | x_j, \boldsymbol{\theta}, \mathbf{y})} \Big|_{\boldsymbol{x}_{-j} = \boldsymbol{\mu}_{-j}}$$



# Wishes and dreams

- ① How can we get a good and cheap approximation  $\tilde{q}(X_j|\boldsymbol{\theta}, \mathbf{y})$  using  $\pi_G(\mathbf{X}|\boldsymbol{\theta}, \mathbf{y})$ ?  
Non-Gaussian? Other family?
- ② How can we remove  $\boldsymbol{\eta}$  from  $\mathbf{X}$  and still produce full posteriors of  $\boldsymbol{\eta}$ ?  
Huge data? Prediction? Stability?



# Gaussian approximation $\pi_G(\boldsymbol{X}|\boldsymbol{\theta}, \boldsymbol{y})$

Laplace method - fit the best Gaussian at the mode of a curve where the variance is derived from the inverse Hessian at the mode.

$$\begin{aligned}\log(\pi(\boldsymbol{X}|\boldsymbol{\theta}, \boldsymbol{y})) &\propto -\frac{1}{2}\boldsymbol{X}^\top \boldsymbol{Q}(\boldsymbol{\theta})\boldsymbol{X} + \sum_{i=1}^n \left( b_i \boldsymbol{X}_i - \frac{1}{2} c_i \boldsymbol{X}_i^2 \right) \\ &= -\frac{1}{2}\boldsymbol{X}^\top (\boldsymbol{Q}(\boldsymbol{\theta}) + \boldsymbol{D})\boldsymbol{X} - \boldsymbol{b}^\top \boldsymbol{X}\end{aligned}$$

hence

$$\boldsymbol{X}|\boldsymbol{\theta}, \boldsymbol{y} \sim N(\boldsymbol{\mu}, (\boldsymbol{Q}(\boldsymbol{\theta}) + \text{diag}(\boldsymbol{c}))^{-1}) \quad (1)$$

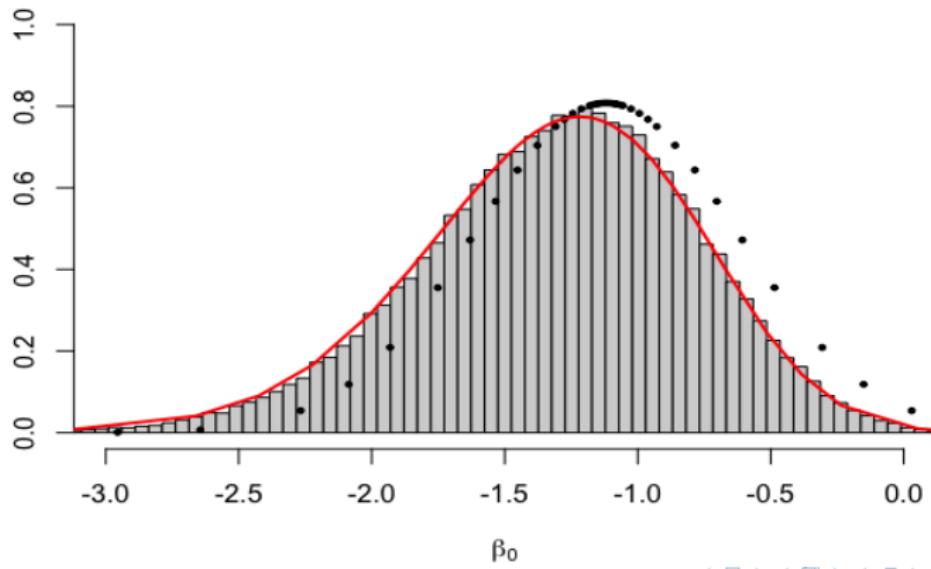
with  $(\boldsymbol{Q}(\boldsymbol{\theta}) + \boldsymbol{D})\boldsymbol{\mu} = \boldsymbol{b}$ .

Can we do better than this?



# Poisson example

$$Y_i | \beta_0 = -1, \beta_1 = -0.5 \sim \text{Poisson}(\exp(\beta_0 + \beta_1 X_i))$$





# Corrected Gaussian approximation

Can we find a better Gaussian approximation (closer to LA) that is cheap and will scale well?

Low-Rank VB correction<sup>1</sup>

---

<sup>1</sup>van Niekerk, J. and Rue, H., 2024. Low-rank Variational Bayes correction to the Laplace method. Journal of Machine Learning Research, 25(62), pp.1-25.



## VB from Zellner (1988)<sup>2</sup>

Based on prior information  $\mathcal{I}$ , data  $\mathbf{y}$  and parameters  $\boldsymbol{\theta}$ , define the following:

- ①  $\pi(\boldsymbol{\theta}|\mathcal{I})$  is the prior model
- ②  $q(\boldsymbol{\theta}|\mathcal{D})$  is the learned model from the prior information and the data where  $\mathcal{D} = \{\mathcal{I}, \mathbf{y}\}$
- ③  $l(\boldsymbol{\theta}|\mathbf{y}) = f(\mathbf{y}|\boldsymbol{\theta})$  is the likelihood
- ④  $p(\mathbf{y}|\mathcal{I})$  is the marginal model for the data where  

$$p(\mathbf{y}|\mathcal{I}) = \int f(\mathbf{y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}|\mathcal{I})d\boldsymbol{\theta}$$

The input information in the learning of  $\boldsymbol{\theta}$  is given by  $\pi(\boldsymbol{\theta}|\mathcal{I})$  and  $l(\boldsymbol{\theta}|\mathbf{y})$ . An information processing rule (IPR) then delivers  $q(\boldsymbol{\theta}|\mathcal{D})$  and  $p(\mathbf{y}|\mathcal{I})$  as output information.

---

<sup>2</sup>Zellner, A., 1988. Optimal information processing and Bayes's theorem. The American Statistician, 42(4), pp.278-280.



# Bayes Rule as an efficient IPR

A stable and efficient IPR would provide the same amount of output information than received through the input information, thus being information conservative. Thus, we learn  $q(\boldsymbol{\theta}|\mathcal{D})$  such that it minimizes

$$\begin{aligned}
 & - \int [\log \pi(\boldsymbol{\theta}|\mathcal{I}) + \log l(\boldsymbol{\theta}|\mathbf{y})] q(\boldsymbol{\theta}|\mathcal{D}) d\boldsymbol{\theta} + \int [\log q(\boldsymbol{\theta}|\mathcal{D}) + \log p(\mathbf{y}|\mathcal{I})] q(\boldsymbol{\theta}|\mathcal{D}) d\boldsymbol{\theta} \\
 & = E_{q(\boldsymbol{\theta}|\mathcal{D})} [-\log l(\boldsymbol{\theta}|\mathbf{y})] + \text{KLD}[q(\boldsymbol{\theta}|\mathcal{D})||\pi(\boldsymbol{\theta}|\mathcal{I})]. \tag{2}
 \end{aligned}$$



# Variational form of Bayes' theorem

Finding the best fit from a certain family  $P(\Theta)$ ,

$$\arg \min_{p \in P(\Theta)} \left( E_{p(\theta)} \left[ - \sum_{i=1}^n \log f(y_i | \theta) \right] + \text{KLD}(p || \pi) \right) \quad (3)$$



# How can we use this?

We apply this to the Gaussian approximation in the denominator.

Recall that  $(Q(\theta) + D)\mu = Q\mu = b$ .

Now let's formulate  $\mu^* = \mu + \delta$ .

$$\arg_{\delta} \min_{p(X|y, \theta)} \left( E_{p(X|y, \theta)} \left[ -\sum_{i=1}^n \log f(y_i | X_i, \theta) \right] + \text{KLD}(p || \pi) \right)$$

where  $X|y, \theta \sim N(\mu + \delta, Q^{-1})$ .

But  $X$  can be very large...



# How can we use this?

We apply this to the Gaussian approximation in the denominator.

Recall that  $(Q(\theta) + D)\mu = Q\mu = b$ .

Now let's formulate  $\mu^* = \mu + \delta$ .

$$\arg_{\delta} \min_{p(X|y, \theta)} \left( E_{p(X|y, \theta)} \left[ - \sum_{i=1}^n \log f(y_i | X_i, \theta) \right] + \text{KLD}(p || \pi) \right)$$

where  $X|y, \theta \sim N(\mu + \delta, Q^{-1})$ .

But  $X$  can be very large...



# How can we use this?

We apply this to the Gaussian approximation in the denominator.

Recall that  $(Q(\theta) + D)\mu = Q\mu = b$ .

Now let's formulate  $\mu^* = \mu + \delta$ .

$$\arg_{\delta} \min_{p(X|y, \theta)} \left( E_{p(X|y, \theta)} \left[ - \sum_{i=1}^n \log f(y_i | X_i, \theta) \right] + \text{KLD}(p || \pi) \right)$$

where  $X|y, \theta \sim N(\mu + \delta, Q^{-1})$ .

But  $X$  can be very large...



# How can we use this?

We apply this to the Gaussian approximation in the denominator.

Recall that  $(Q(\boldsymbol{\theta}) + D)\boldsymbol{\mu} = Q\boldsymbol{\mu} = \mathbf{b}$ .

Now let's formulate  $\boldsymbol{\mu}^* = \boldsymbol{\mu} + \boldsymbol{\delta}$ .

$$\arg_{\boldsymbol{\delta}} \min_{p(\mathbf{X}|\mathbf{y}, \boldsymbol{\theta})} \left( E_{p(\mathbf{X}|\mathbf{y}, \boldsymbol{\theta})} \left[ - \sum_{i=1}^n \log f(y_i | X_i, \boldsymbol{\theta}) \right] + \text{KLD}(p || \pi) \right)$$

where  $\mathbf{X}|\mathbf{y}, \boldsymbol{\theta} \sim N(\boldsymbol{\mu} + \boldsymbol{\delta}, Q^{-1})$ .

But  $\mathbf{X}$  can be very large...



# Implicit mean correction

Recall that  $\mathbf{Q}\boldsymbol{\mu} = \mathbf{b}$ .

Now let's formulate  $\mathbf{Q}\boldsymbol{\mu}^* = \mathbf{b} + \boldsymbol{\lambda}$ , so that

$$\boldsymbol{\mu}^* = \boldsymbol{\mu} + \mathbf{M}\boldsymbol{\lambda}$$

So now we solve for,

$$\arg_{\boldsymbol{\lambda}} \min_{p(\mathbf{X}|\mathbf{y}, \boldsymbol{\theta})} \left( E_{p(\mathbf{X}|\mathbf{y}, \boldsymbol{\theta})} \left[ -\sum_{i=1}^n \log f(y_i | \mathbf{X}_i, \boldsymbol{\theta}) \right] + \text{KLD}(p || \pi) \right)$$

where  $\mathbf{X}|\mathbf{y}, \boldsymbol{\theta} \sim N(\boldsymbol{\mu}^*, \mathbf{Q}^{-1})$ .

Low-rank correction → Only correct some  $b$ 's, change to all  $\mu$ 's.



# Implicit mean correction

Recall that  $\mathbf{Q}\boldsymbol{\mu} = \mathbf{b}$ .

Now let's formulate  $\mathbf{Q}\boldsymbol{\mu}^* = \mathbf{b} + \boldsymbol{\lambda}$ , so that

$$\boldsymbol{\mu}^* = \boldsymbol{\mu} + \mathbf{M}\boldsymbol{\lambda}$$

So now we solve for,

$$\arg_{\boldsymbol{\lambda}} \min_{p(\mathbf{X}|\mathbf{y}, \boldsymbol{\theta})} \left( E_{p(\mathbf{X}|\mathbf{y}, \boldsymbol{\theta})} \left[ - \sum_{i=1}^n \log f(y_i | \mathbf{X}_i, \boldsymbol{\theta}) \right] + \text{KLD}(p || \pi) \right)$$

where  $\mathbf{X}|\mathbf{y}, \boldsymbol{\theta} \sim N(\boldsymbol{\mu}^*, \mathbf{Q}^{-1})$ .

Low-rank correction → Only correct some  $b$ 's, change to all  $\mu$ 's.



# Implicit mean correction

Recall that  $\mathbf{Q}\boldsymbol{\mu} = \mathbf{b}$ .

Now let's formulate  $\mathbf{Q}\boldsymbol{\mu}^* = \mathbf{b} + \boldsymbol{\lambda}$ , so that

$$\boldsymbol{\mu}^* = \boldsymbol{\mu} + \mathbf{M}\boldsymbol{\lambda}$$

So now we solve for,

$$\arg_{\boldsymbol{\lambda}} \min_{p(\mathbf{X}|\mathbf{y}, \boldsymbol{\theta})} \left( E_{p(\mathbf{X}|\mathbf{y}, \boldsymbol{\theta})} \left[ - \sum_{i=1}^n \log f(y_i | \mathbf{X}_i, \boldsymbol{\theta}) \right] + \text{KLD}(p || \pi) \right)$$

where  $\mathbf{X}|\mathbf{y}, \boldsymbol{\theta} \sim N(\boldsymbol{\mu}^*, \mathbf{Q}^{-1})$ .

Low-rank correction → Only correct some  $b$ 's, change to all  $\mu$ 's.



# Implicit mean correction

Recall that  $\mathbf{Q}\boldsymbol{\mu} = \mathbf{b}$ .

Now let's formulate  $\mathbf{Q}\boldsymbol{\mu}^* = \mathbf{b} + \boldsymbol{\lambda}$ , so that

$$\boldsymbol{\mu}^* = \boldsymbol{\mu} + \mathbf{M}\boldsymbol{\lambda}$$

So now we solve for,

$$\arg_{\boldsymbol{\lambda}} \min_{p(\mathbf{X}|\mathbf{y}, \boldsymbol{\theta})} \left( E_{p(\mathbf{X}|\mathbf{y}, \boldsymbol{\theta})} \left[ - \sum_{i=1}^n \log f(y_i | \mathbf{X}_i, \boldsymbol{\theta}) \right] + \text{KLD}(p || \pi) \right)$$

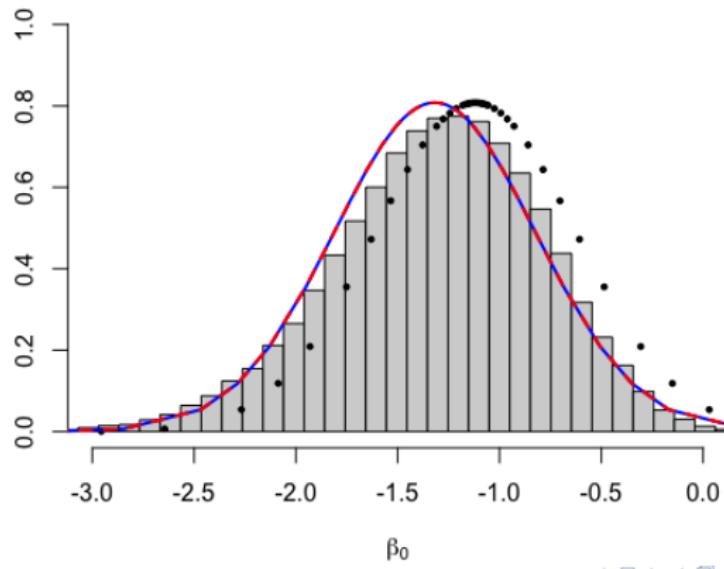
where  $\mathbf{X}|\mathbf{y}, \boldsymbol{\theta} \sim N(\boldsymbol{\mu}^*, \mathbf{Q}^{-1})$ .

Low-rank correction → Only correct some  $b$ 's, change to all  $\mu$ 's.



## Example (small data)

$$Y_i | \beta_0 = -1, \beta_1 = -0.5 \sim \text{Poisson}(\exp(\beta_0 + \beta_1 X_i))$$





# Results of mean correction

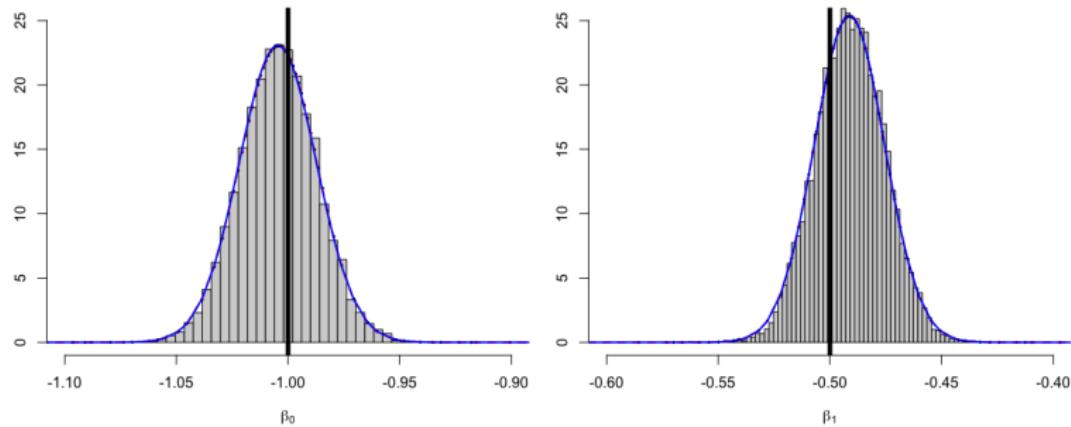
	GA	INLA 1.0	VBC	MCMC
$\beta_0$	-1.119	-1.317	-1.316	-1.302
$\beta_1$	-0.361	-0.401	-0.401	-0.391

Table: Posterior means from the Gaussian method, INLA, VBC and MCMC



## Example (large data)

$$Y_i | \beta_0 = -1, \beta_1 = -0.5 \sim \text{Poisson}(\exp(\beta_0 + \beta_1 X_i))$$



**Figure:** Marginal posterior of  $\beta_0$  (center) and  $\beta_1$  (right) from the Laplace method (points), VBC (solid line) and INLA (broken line) approximations



# Results

	GA	INLA 1.0	VBC	MCMC
$\beta_0$	-1.004	-1.005	-1.005	-1.005
$\beta_1$	-0.491	-0.492	-0.492	-0.492
Time (s)	6.346	37.168	7.344	5779.42

Table: Posterior means from the Laplace method, INLA 1.0, VBC and MCMC



## Conditional posterior of $\eta_i$

In order to calculate  $\tilde{\pi}(\eta_i|\mathbf{y})$ , we first calculate  $\tilde{\pi}(\eta_i|\boldsymbol{\theta}, \mathbf{y})$ . We postulate a Gaussian density for  $\eta_i|\boldsymbol{\theta}, \mathbf{y}$  such that  $\tilde{\pi}(\eta_i|\boldsymbol{\theta}, \mathbf{y}) = \pi_G(\eta_i|\boldsymbol{\theta}, \mathbf{y})$ , with mean

$$E(\boldsymbol{\eta}|\boldsymbol{\theta}, \mathbf{y}) = \mathbf{A}E(\mathbf{X}|\boldsymbol{\theta}, \mathbf{y}) = \mathbf{A}\boldsymbol{\mu}(\boldsymbol{\theta})$$

and covariance matrix

$$\text{Cov}(\boldsymbol{\eta}|\boldsymbol{\theta}, \mathbf{y}) = \mathbf{A}\text{Cov}(\mathbf{X}|\boldsymbol{\theta}, \mathbf{y})\mathbf{A}^\top,$$



# VB corrected marginal posterior of $\eta_i$

$$\begin{aligned}\eta_j | \boldsymbol{\theta}, \mathbf{y} &\sim N(\mu_j(\boldsymbol{\theta}), \sigma_j^2(\boldsymbol{\theta})) \\ \mu_j(\boldsymbol{\theta}) &= (\mathbf{A}\boldsymbol{\mu}^*(\boldsymbol{\theta}))_j \\ \tilde{\pi}(\eta_j | \mathbf{y}) &\approx \sum_{k=1}^K \pi_G(\eta_j | \boldsymbol{\theta}_k, \mathbf{y}) \tilde{\pi}(\boldsymbol{\theta}_k | \mathbf{y}) \delta_k.\end{aligned}$$



# Universal tools in INLA

- Model selection metrics - WAIC, DIC
- Cross validation (1 and group) and model-based clustering
- Prediction of unobserved areas or new profiles
- Mean or quantile models
- Joint models
- Multiple imputation
- Coregionalization models
- etc..... ask at <https://groups.google.com/g/r-inla-discussion-group?pli=1> or  
e-mail [help@r-inla.org](mailto:help@r-inla.org)



## Tokyo example

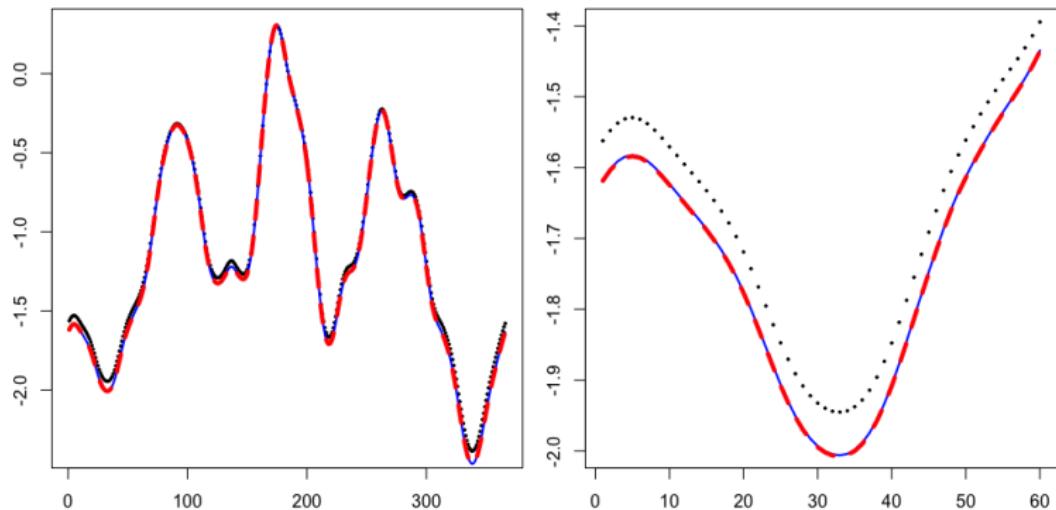
The Tokyo dataset in the R INLA library contains information on the number of times the daily rainfall measurements in Tokyo was more than 1mm on a specific day  $t$  for two consecutive years. In order to model the annual rainfall pattern, a stochastic spline model with fixed precision is used to smooth the data.

$$\begin{aligned}y_i | \mathcal{X} &\sim \text{Bin}\left(n_i, p_i = \frac{\exp(\alpha_i)}{1 + \exp(\alpha_i)}\right) \\(\alpha_{i+1} - 2\alpha_i + \alpha_{i-1}) | \tau &\stackrel{\text{iid}}{\sim} N(0, \tau^{-1}),\end{aligned}$$

where  $i = 1, 2, \dots, 366$  on a torus, and  $n_{60} = 1$  else  $n_i = 2$ .



# Results



**Figure:** Posterior mean of  $\alpha$  (left) (zoomed for the first two months (right)) from the Laplace method (points), VBC (solid line) and INLA (broken line)



# Cox proportional hazards model

We simulate survival data for  $n$  patients using the following very simple Cox proportional hazards model

$$h_i(t) = h_0(t) \exp(\beta x_i) = 1.2t^{0.2} \exp(0.1x_i), \quad i = 1, 2, \dots, n,$$

where  $x$  is a scaled and centered continuous covariate, and the baseline hazard,  $h_0(t)$  is estimated using a scaled random walk order one model with 50 bins. We also consider four different values of  $n$  which are  $n = 10^2$ , to  $10^5$ .



# Cox proportional hazards model

$n$	Augmented size	classic INLA (s)	modern INLA (s)
$10^2$	1 327	1.6	0.1
$10^3$	12 657	1.3	0.4
$10^4$	131 807	10.2	2.3
$10^5$	1 302 413	113.3	22.5

**Table:** Results from simulation of Cox proportional hazards model



## Flexible Besag model<sup>3</sup>

Instead of one precision for the entire area, we define multiple precision parameters,  $\tau_1, \tau_2, \dots, \tau_P$ , to account for covariance non-stationarity. The conditional density for the spatial effect of area  $i$  is

$$x_i | \mathbf{x}_{-i}, \tau_1, \dots, \tau_P \sim N\left(\frac{1}{2} \sum_{\substack{i \text{ in sub-region } k \\ j \text{ in sub-region } l \\ i \sim j}} (\tau_l + \tau_k) \tau_{x_i}^{-1} x_j, \tau_{x_i}^{-1}\right),$$

and

$$\tau_{x_i} = \frac{1}{2} \left( n_i \tau_k + \sum_l n_{il} \tau_l \right).$$

---

<sup>3</sup>Abdul-Fattah, E., Krainski, E., Van Niekerk, J. and Rue, H., 2024. Non-stationary Bayesian spatial model for disease mapping based on sub-regions. Statistical Methods in Medical Research, p.09622802241244613.



# Contraction prior: Non-stationary → stationary

The joint PC prior for  $\boldsymbol{\theta} = \log \boldsymbol{\tau}$  can be derived as a convolution of the PC prior for  $\boldsymbol{\tau}$  from the Besag model, as follows

$$\pi(\boldsymbol{\theta}) = 2^{-(P+2)/2} \pi^{-P/2} \lambda \sigma^{-P} \exp\left(-\frac{1}{2}(\boldsymbol{\theta} - \bar{\boldsymbol{\theta}})^T \tilde{\boldsymbol{\Sigma}}^{-1} (\boldsymbol{\theta} - \bar{\boldsymbol{\theta}}) - \bar{\boldsymbol{\theta}}/2 - \lambda e^{-\bar{\boldsymbol{\theta}}/2}\right),$$

This prior contracts

$$\tau_1, \tau_2, \dots, \tau_P \rightarrow \tau$$



# Dengue risk in Brazil

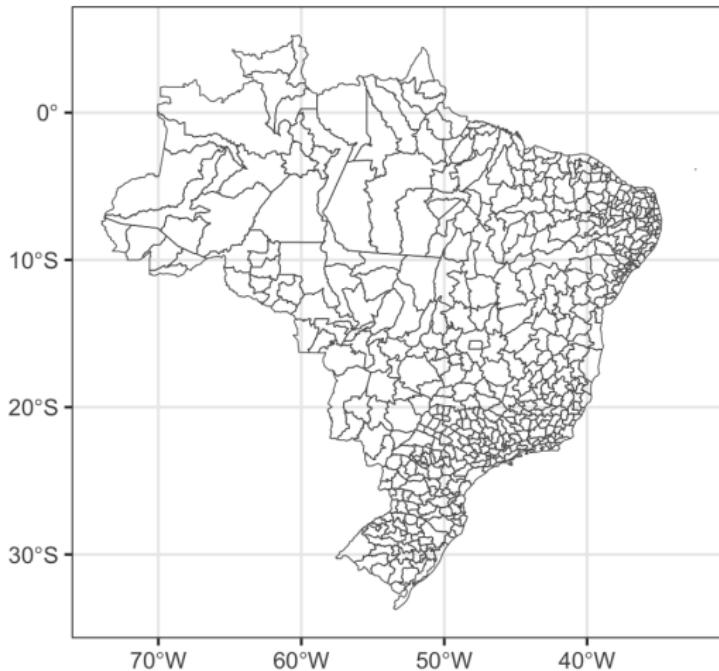
We analyze the effects of hydrometeorological hazards on dengue risk in Brazil. To test the spatial variations in the spread of the virus in different sub-regions of Brazil, we fit dengue counts with a Poisson regression model as follows,

$$\mathbf{y} \sim \text{Poisson}(Ee^{\boldsymbol{\eta}}), \quad \boldsymbol{\eta} = \mathbf{1}^T \boldsymbol{\mu} + \boldsymbol{\alpha}$$

where  $\mathbf{y}$  is the observed counts in November of dengue cases,  $E$  is the expected number of counts,  $\boldsymbol{\eta}$  is the linear predictor,  $\boldsymbol{\mu}$  is the overall intercept, and  $\boldsymbol{\alpha}$  is the Besag or flexible Besag model over space.

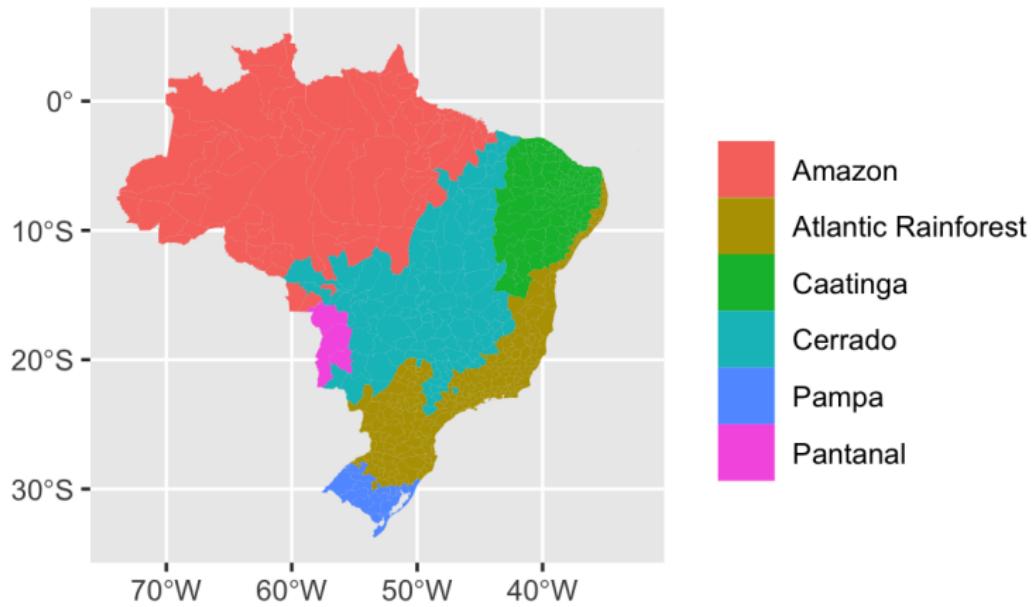


# Dengue risk in Brazil



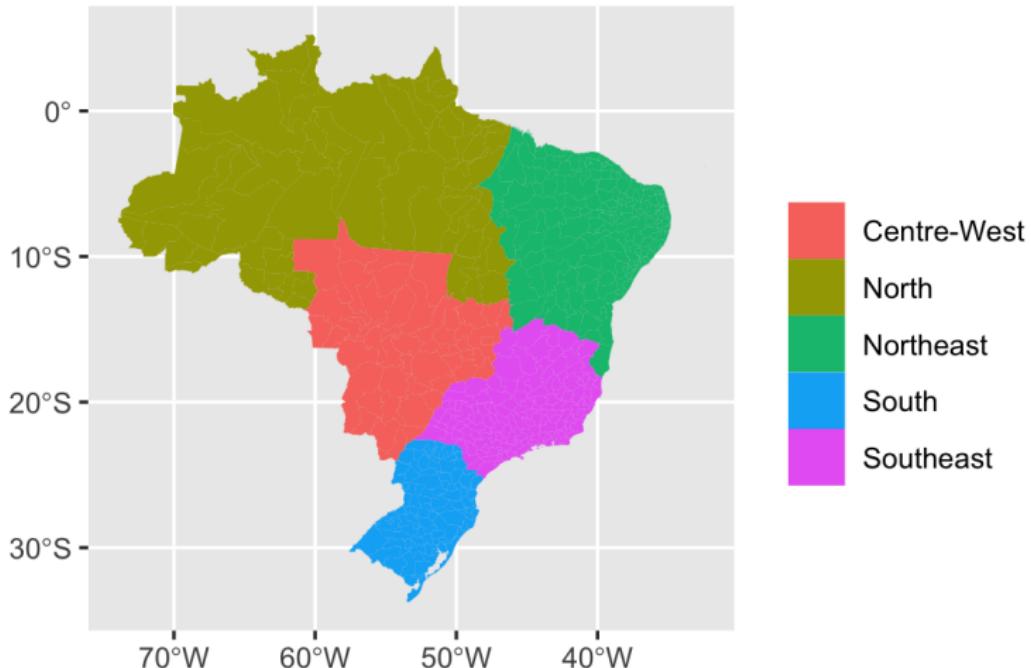


# Dengue risk in Brazil





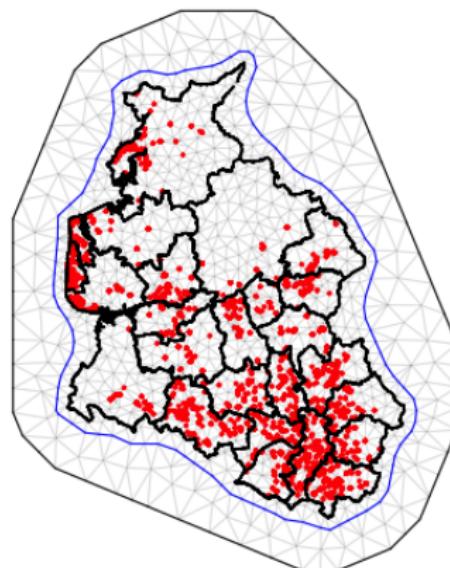
# Dengue risk in Brazil





## Spatial survival example

Consider the R dataset `Leuk` that features the survival times of 1043 patients with acute myeloid leukemia (AML) in Northwest England between 1982 to 1998.





# Cox spatial model

$$h(t, s) = h_0(t) \exp(\beta \mathbf{X} + \mathbf{u}(s)),$$

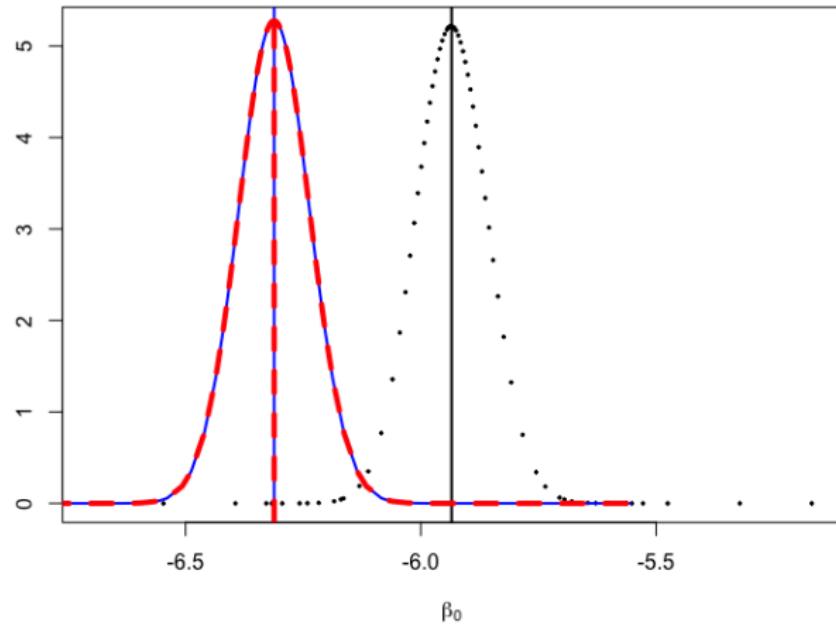
with

$$\eta_i(s) = \beta_0 + \beta_1 \text{Age}_i + \beta_2 \text{WBC}_i + \beta_3 \text{TPI}_i + u(s).$$

which implies a latent field of size  $m = 39158$ .



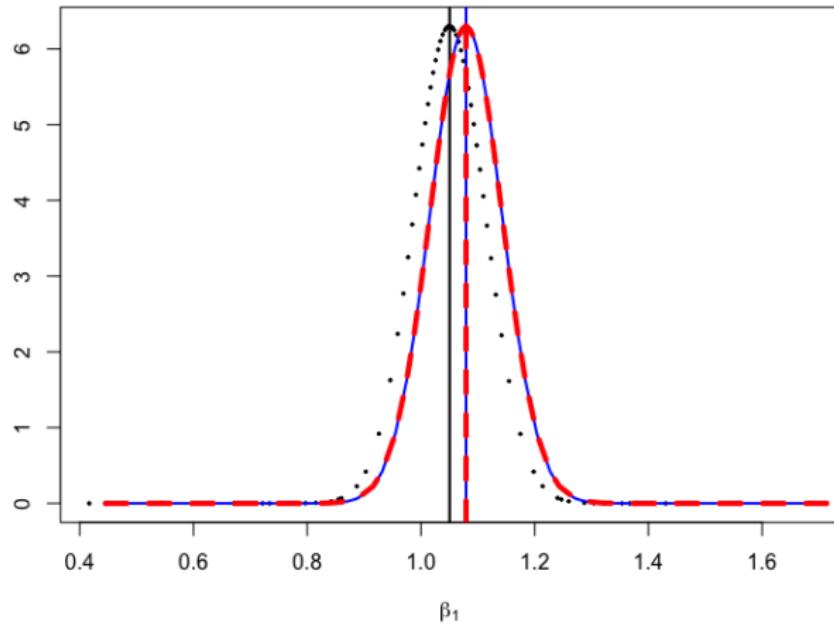
# Results



**Figure:** Marginal posteriors from the Laplace method (points), VBC (solid line) and INLA (broken line)



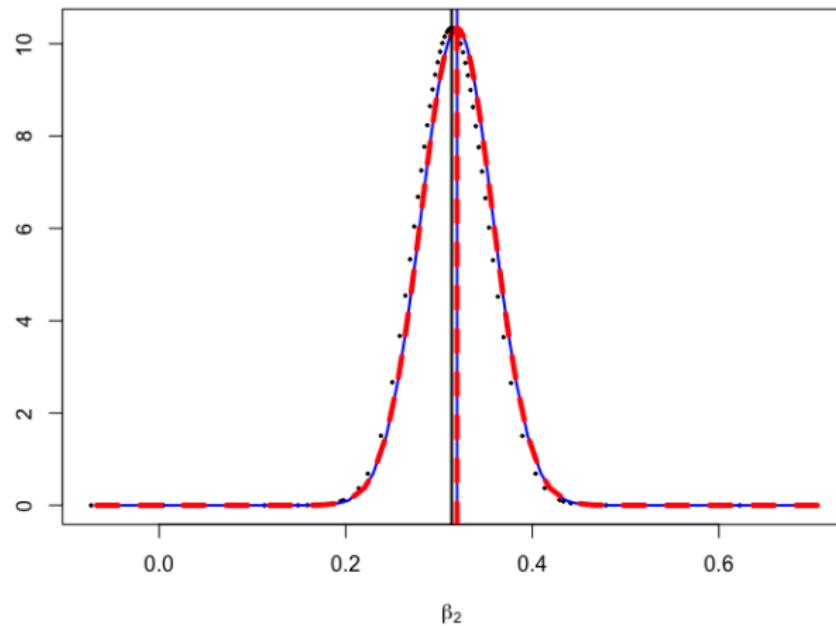
# Results



**Figure:** Marginal posteriors from the Laplace method (points), VBC (solid line) and INLA (broken line)



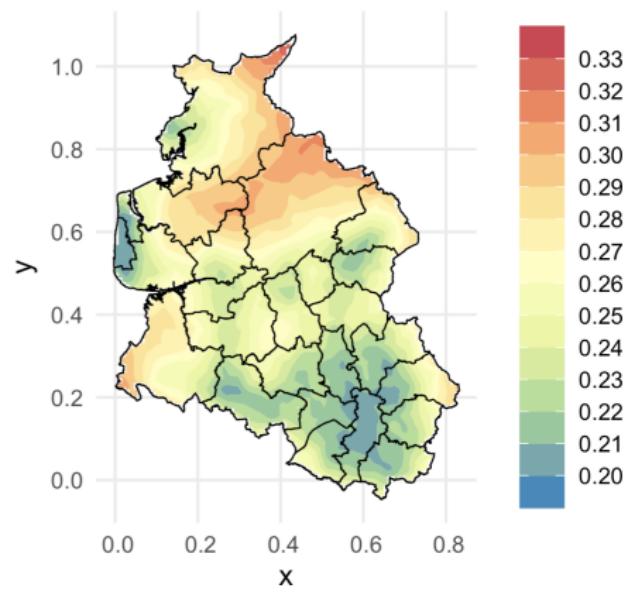
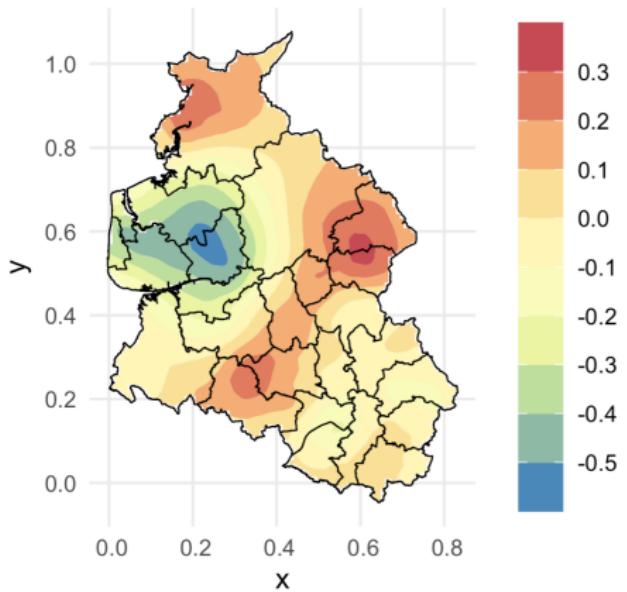
# Results



**Figure:** Marginal posteriors from the Laplace method (points), VBC (solid line) and INLA (broken line)



# Matern field





# The Matérn's SPDE

- Whittle (1954), Whittle (1963):

- Fields with Matérn covariance are solutions to the following Stochastic Partial Differential Equation (SPDE)

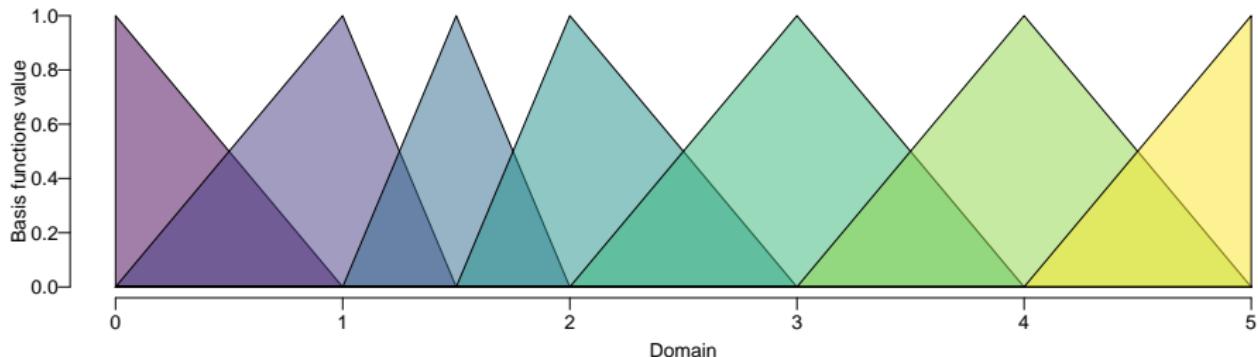
$$\tau(\kappa^2 - \Delta)^{\alpha/2} u(\mathbf{s}) = \mathcal{W}(\mathbf{s})$$

- $\kappa > 0$ : scale parameter
- $\alpha = \nu + d/2$ : smoothness
- $\Delta$  is the Laplacian

$$\Delta = \sum_{i=1}^d \frac{\partial^2}{\partial s_i^2}$$



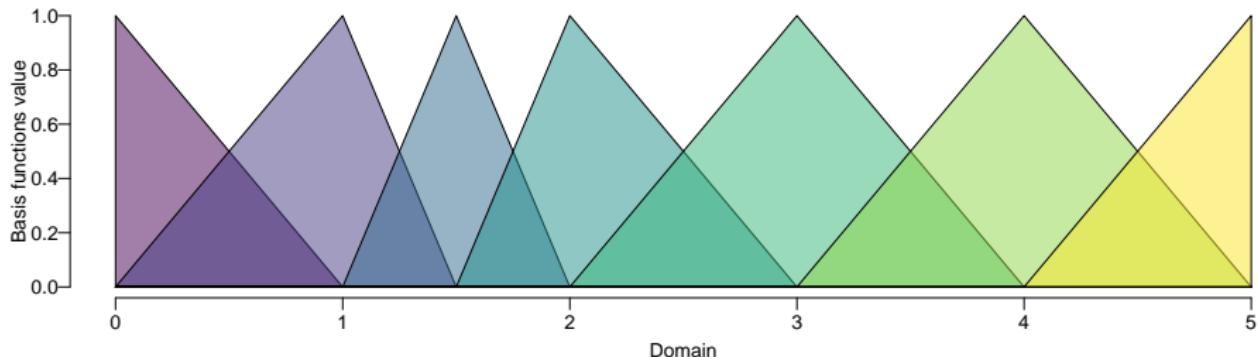
# How to solve the SPDE? FEM



- $u(\mathbf{s}) \approx \sum_{k=1}^m \psi_k(\mathbf{s}) u_k = \mathbf{A}(\mathbf{s}, \mathbf{s}_0) u(\mathbf{s}_0),$ 
  - $\psi_k$ : basis functions evaluated at data locations  $\mathbf{s}$
  - $u_k$ : the process at the discretization points  $\mathbf{s}_0$



# How to solve the SPDE? FEM



- $u(\mathbf{s}) \approx \sum_{k=1}^m \psi_k(\mathbf{s}) u_k = \mathbf{A}(\mathbf{s}, \mathbf{s}_0) u(\mathbf{s}_0)$ ,
  - $\psi_k$ : basis functions evaluated at data locations  $\mathbf{s}$
  - $u_k$ : the process at the discretization points  $\mathbf{s}_0$



# Lindgren, Rue, and Lindström (2011)<sup>4</sup> I

- Discretization

- sparse precision matrix:
- $\mathbf{Q}_\alpha(\tau, \kappa)$ , for  $\alpha \in \{1, 2, \dots\}$ .

- $\alpha$

- $\alpha = 1: \tau \mathbf{K}_1(\kappa) = \tau(\kappa^2 \mathbf{C} + \mathbf{G})$
- $\alpha = 2: \tau(\kappa^4 \mathbf{C} + 2\kappa^2 \mathbf{G} + \mathbf{G}\mathbf{C}^{-1}\mathbf{G})$
- $\alpha = 2, 3, 4, \dots: \tau \mathbf{K}_1(\kappa) \mathbf{C}^{-1} \mathbf{K}_{\alpha-2}(\kappa) \mathbf{C}^{-1} \mathbf{K}_1(\kappa)$

Equivalent models (discretized): Whittle (1954), Besag (1974), Besag (1981), Besag and Kooperberg (1995) and Besag and Mondal (2005).

- $\alpha = 1: \tau \mathbf{K}_1(\kappa) = \tau(\kappa^2 \mathbf{C} + \mathbf{G})$



# Lindgren, Rue, and Lindström (2011)<sup>5</sup> II

- $d=1, u_1, u_2, \dots, u_n$ , two neighbours

$$\tau^2 \begin{bmatrix} 1 + \kappa^2 & -1 & & \\ -1 & 2 + \kappa^2 & -1 & \\ & \ddots & & \\ & -1 & 2 + \kappa^2 & -1 \\ & & -1 & 1 + \kappa^2 \end{bmatrix}$$

- $d = 2, \mathbf{C} = \mathbf{I}, \mathbf{G} = \text{Laplacian}$  (4 neighbours)

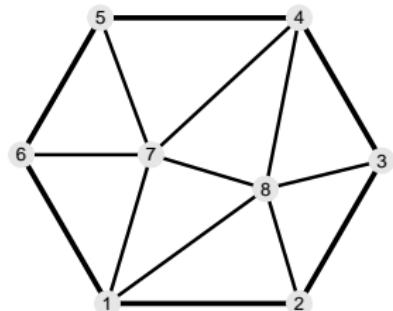
<sup>4</sup>Lindgren, F., Rue, H. and Lindström, J., 2011. An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. Journal of the Royal Statistical Society Series B: Statistical Methodology, 73(4), pp.423-498.

<sup>5</sup>Lindgren, F., Rue, H. and Lindström, J., 2011. An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. Journal of the Royal Statistical Society Series B: Statistical Methodology, 73(4), pp.423-498.

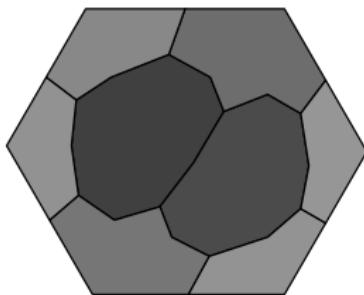


# Piecewise linear basis, FEM matrices

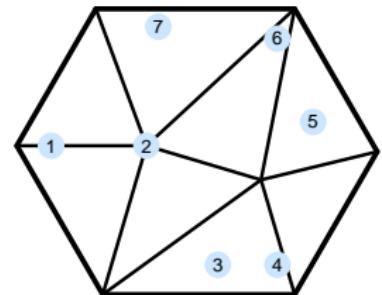
Mesh nodes



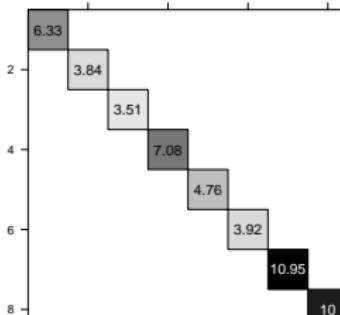
Dual mesh



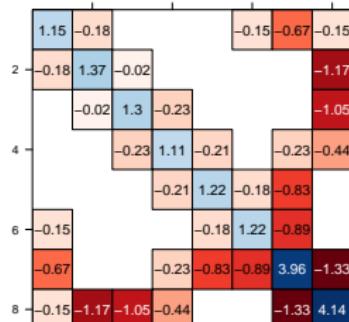
Data locations



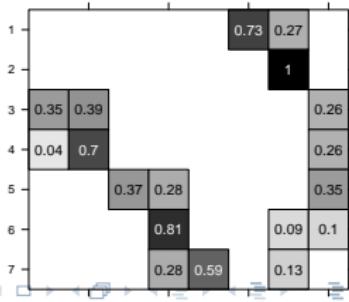
C



G

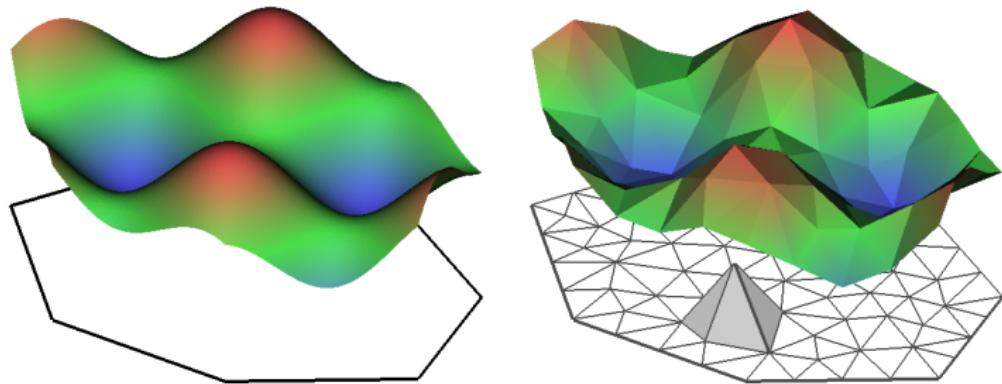


A





# FEM in 3D

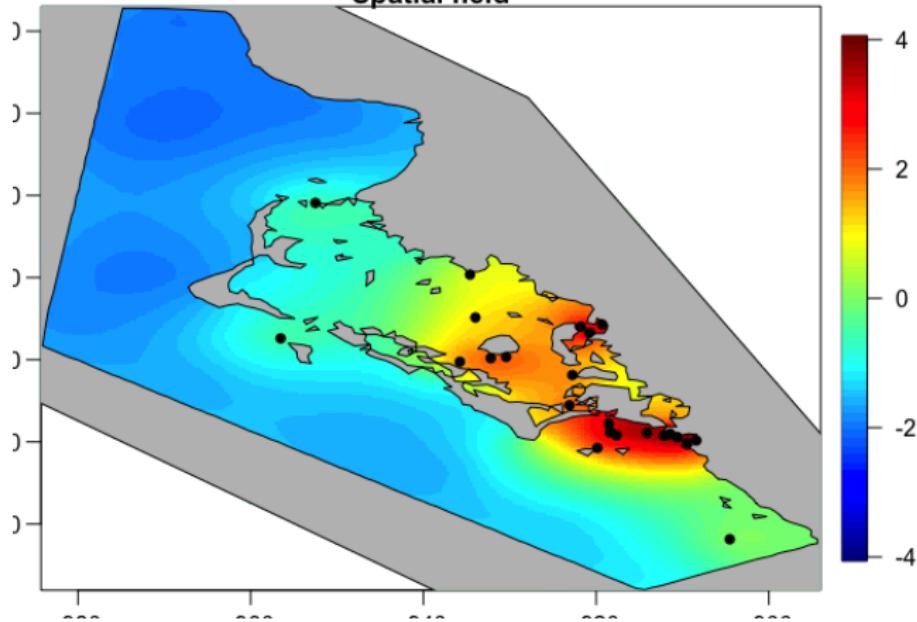




# Non-stationary Matern field based on physical barriers

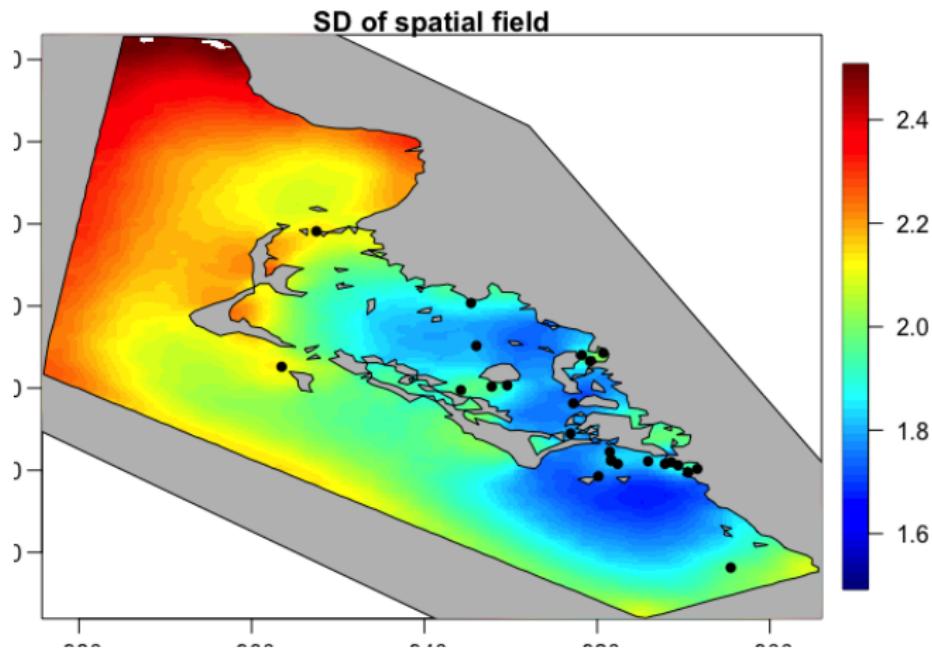
Now the distance is not Euclidean... Construct the mesh with boundaries

Spatial field





# Non-stationary Matern field based on physical barriers





## cs-fMRI model<sup>6</sup>

Functional magnetic resonance imaging (fMRI) is a noninvasive neuro-imaging technique used to localize regions of specific brain activity during certain tasks.

For  $T$  timepoints and  $N$  vertices per hemisphere resulting in data  $\mathbf{y}_{TN \times 1}$  with the latent Gaussian model as follows:

$$\begin{aligned}\mathbf{y}|\boldsymbol{\beta}, \mathbf{b}, \boldsymbol{\theta} &\sim N(\boldsymbol{\mu}_y, \mathbf{V}), \quad \boldsymbol{\mu}_y = \sum_{k=0}^K \mathbf{X}_k \boldsymbol{\beta}_k + \sum_{j=1}^J \mathbf{Z}_j \mathbf{b}_j \\ \boldsymbol{\beta}_k &= \boldsymbol{\Psi}_k \mathbf{w}_k \quad (\text{SPDE prior on } \boldsymbol{\beta}_k) \\ \mathbf{w}_k | \boldsymbol{\theta} &\sim N(\mathbf{0}, \mathbf{Q}_{\tau_k, \kappa_k}^{-1}) \\ \mathbf{b}_j &\sim N(\mathbf{0}, \delta \mathbf{I}) \quad (\text{Diffuse priors for } \mathbf{b}_j) \\ \boldsymbol{\theta} &\sim \pi(\boldsymbol{\theta}),\end{aligned}$$

where we have  $K$  task signals and  $J$  nuisance signals.

<sup>6</sup>Van Niekerk, J., Krainski, E., Rustand, D. and Rue, H., 2023. A new avenue for Bayesian inference with INLA. *Computational Statistics & Data Analysis*, 181, p.107602.



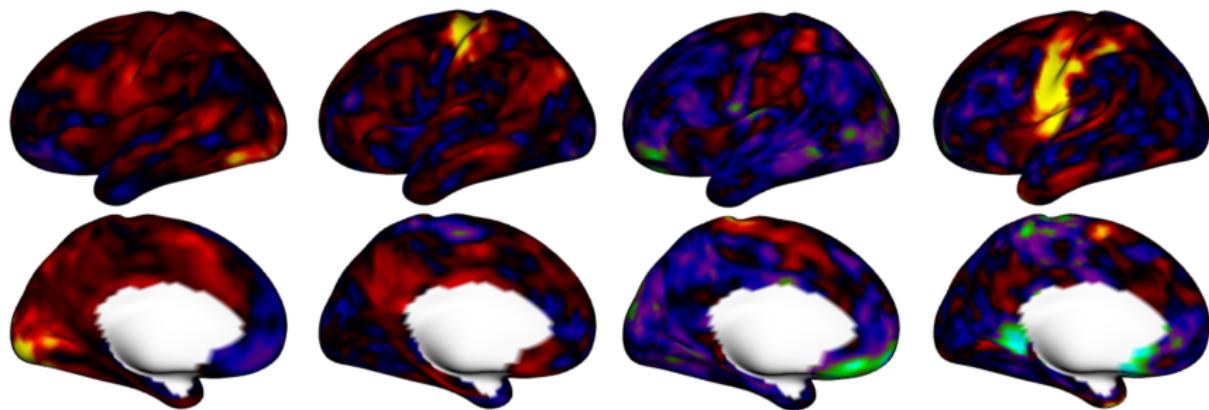
## cs-fMRI model

The data consists of a 3.5-min fMRI for each subject, consisting of 284 volumes, where each subject performs 5 different motor tasks interceded with a 3 second visual cue. Each hemisphere of the brain contained 32492 surface vertices. From these, 5000 are resampled to use for the analysis. This results in a response data vector  $\mathbf{y}$  of size **2 523 624**, with an SPDE model defined on a mesh with 8795 triangles.

The inference based on the modern formulation of INLA was computed in 148 seconds.



# cs-fMRI model



**Figure:** Activation areas for the different tasks in the left hemisphere - visual cue, right hand motor, right foot motor, tongue motor task (from left to right)



## Further details

[www.r-inla.org](http://www.r-inla.org)

New default setting in INLA (VB) (previously `inla.mode = "experimental"`)

- INLA can fit many different statistical models and complex models can be built using multiple "building blocks"/random effects.
- Remove the linear predictors from the latent field → accurate posterior inference with VB correction (I - VB - LA)
- New applications that aren't feasible with INLA 1.0

-  **Gaedke-Merzhäuser, L., van Niekerk, J., Schenk, O., and Rue, H. (2023).**  
Parallelized integrated nested Laplace approximations for fast Bayesian inference.  
*Statistics and Computing*, 33(1):25.
-  **Rue, H., Martino, S., and Chopin, N. (2009).**  
Approximate Bayesian inference for latent Gaussian models by using integrated nested laplace approximations.  
*Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 71(2):319–392.
-  **Van Niekerk, J., Krainski, E., Rustand, D., and Rue, H. (2023).**  
A new avenue for Bayesian inference with INLA.  
*Computational Statistics & Data Analysis*, 181:107692.
-  **van Niekerk, J. and Rue, H. (2024).**  
Low-rank variational Bayes correction to the Laplace method.  
*Journal of Machine Learning Research*, 25(62):1–25.



شكراً • Thank you



جامعة الملك عبد الله  
للغعلوم والتكنولوجية  
King Abdullah University of  
Science and Technology