

Approximate Bayesian inference with INLA

Janet van Niekerk

Ethiopian Public Health Institute

King Abdullah University of
Science and Technology



جامعة الملك عبد الله
لعلوم والتكنولوجيا



Ethiopian Public Health Institute
ኢትዮጵያ የአብዛኛዎች ሥርዓት አገልግሎት

Repository for the workshop



<https://github.com/JanetVN1201/EPHI>

Outline I

- 1 Bayesian inference
- 2 Latent Gaussian models
- 3 Posterior inference with INLA
- 4 R-INLA
- 5 Priors
- 6 Examples
 - AML survival (using Matern field)
 - Adherence study (Quantile joint model)
 - Dengue risk in Brazil (Non-stationary disease mapping)
 - Dementia study (3D Spatial model)

BayesComp group at KAUST



Survival analysis



Survival analysis

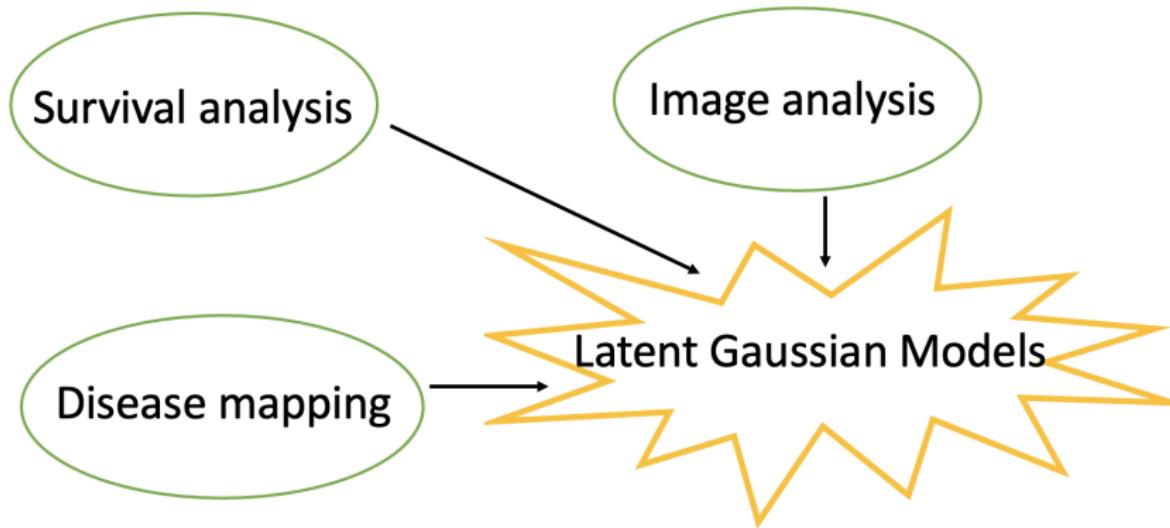
Image analysis



Survival analysis

Image analysis

Disease mapping



Bayesian inference



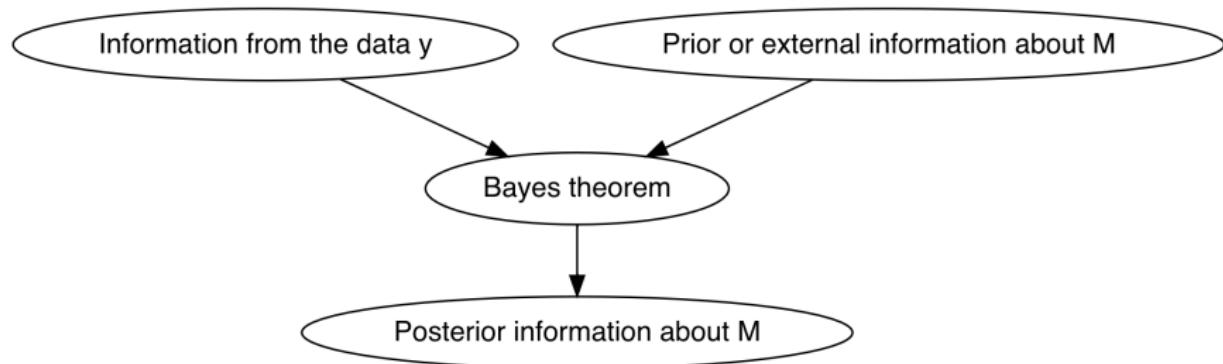
In GLM usually we model the mean with a linear model

$$E(y) = g^{-1}(\omega^\top Z)$$

Bayesian learning



Conceptually we can illustrate Bayesian statistics with the following diagram.





Bayes' theorem

- Exact Bayesian inference requires computing posterior:

$$p(\omega | y) = \frac{p(y | \omega)p(\omega)}{p(y)}.$$

- Problem: marginal likelihood

$$p(y) = \int p(y | \omega)p(\omega) d\omega$$

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.



What is INLA?

INLA - Integrated Nested Laplace Approximations

- Deterministic approximations instead of sampling
- LGM - Latent Gaussian models
- Three internal strategies - Gaussian, simplified Laplace, Laplace (pre 2021)
- R package "INLA"

Now there is a new default strategy combining Laplace approximations with Variational Bayes.¹ (2021+)

¹van 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.

INLA versus MCMC



For small models and data:

	INLA	MCMC/HMC
Method	Deterministic (Mathematical)	Sampling-based
Memory cost	Low	Low
Computational cost	Low	Low

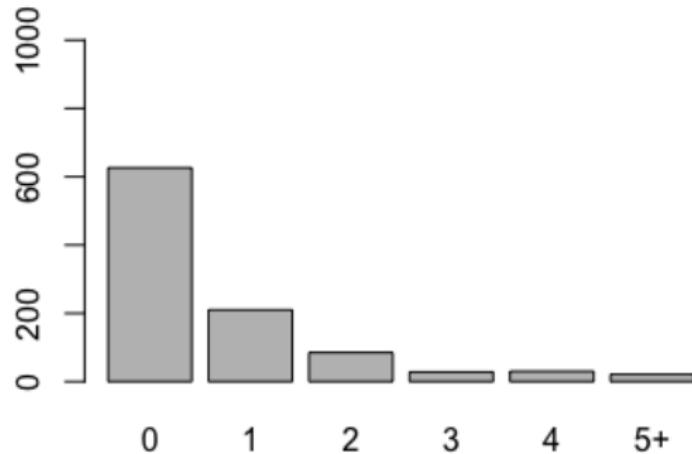
For large models and/or data:

	INLA	MCMC/HMC
Method	Deterministic (Mathematical)	Sampling-based
Memory cost	Low	High
Computational cost	Low	Very high

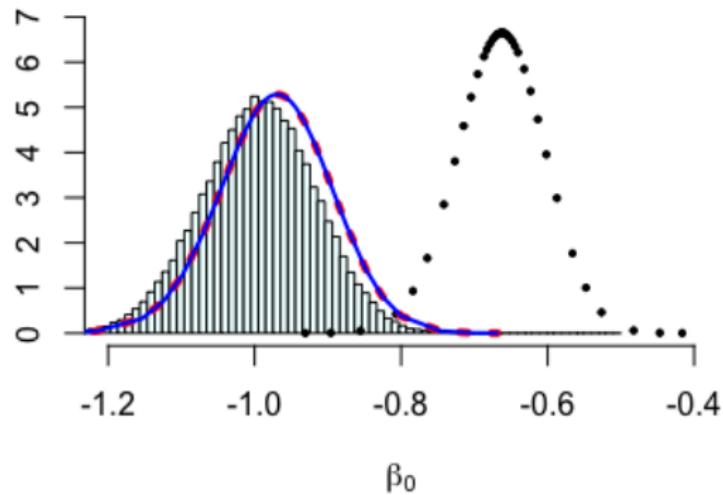
Example - INLA and MCMC I



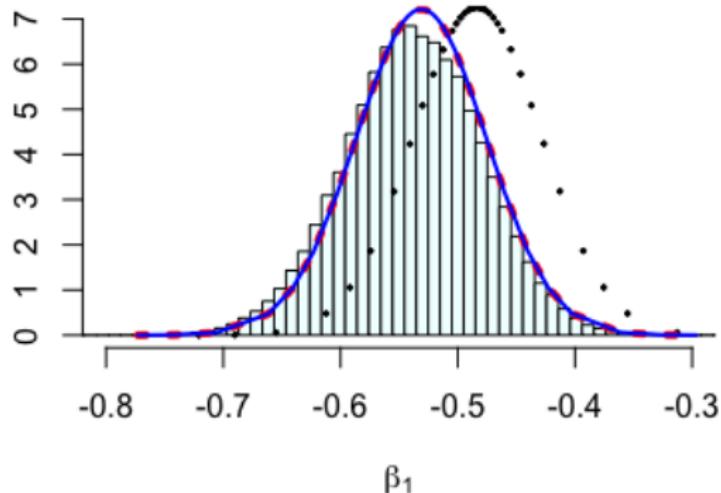
$$y_i \sim \text{Poisson}(\exp(\eta_i)), \quad \eta_i = \beta_0 + \beta_1 x_i + u_i$$



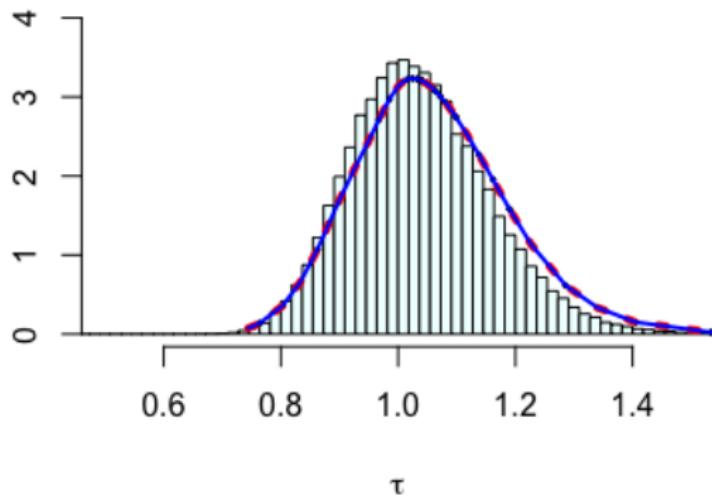
Example - INLA and MCMC II



Example - INLA and MCMC III



Example - INLA and MCMC IV



Example - INLA and MCMC V



	INLA	MCMC
β_0	-0.972	-0.934
β_1	-0.531	-0.529
τ	1.056	1.037
Time(s)	5.718	207.445

Why is INLA so accurate and so fast?



- LGM structure
- Sparse precision matrix
- Specialized matrix algebra for sparse matrices
- NEW: VB (low-rank) correction

Use precision matrix instead of covariance matrix → natural occurrence

How common are sparse precision matrices?



Consider an AR(1) model..

AR(1) example



Ethiopian Public Health Institute
ኢትዮጵያ የክብረተሰቦች ሚኒስቴር

AR(1) example



Ethiopian Public Health Institute
ኢትዮጵያ የክብረተሰቦች ሚኒስቴር

But what about public health models like disease mapping?
Can we use INLA to perform Bayesian inference of such models?

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

$$\boldsymbol{\eta}_n = \boldsymbol{\beta}_0 + \mathbf{Z}_\beta \boldsymbol{\beta} + \sum f^k(\mathbf{Z}_f) = \mathbf{A}\mathbf{X}$$

The inferential aim is to estimate the latent field $\mathbf{X}_m = \{\boldsymbol{\beta}_0, \boldsymbol{\beta}, \mathbf{f}\}$, 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.



INLA is designed to work for LGM's

- ① Data y_i with some likelihood $L(\mathbf{X}, \boldsymbol{\theta} | \mathbf{y}) = \prod_{i=1}^n f(y_i | \eta_i = h(\mathbf{X}, \boldsymbol{\theta}))$
- ② $\eta_i = h(\beta Z + \mathbf{u}(\boldsymbol{\theta}) \mathbf{A})$
- ③ $\mathbf{X} | \boldsymbol{\theta} \sim N(\mathbf{0}, Q(\boldsymbol{\theta})^{-1})$
- ④ $\boldsymbol{\theta} \sim \text{hyperprior}$



Posterior approximations by INLA

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

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



Modern INLA

The Gaussian approximation $\pi_G(\mathbf{X}|\boldsymbol{\theta}, \mathbf{y})$ to $\pi(\mathbf{X}|\boldsymbol{\theta}, \mathbf{y})$ is calculated from a second order expansion of the likelihood around the mode of $\pi(\mathbf{X}|\boldsymbol{\theta}, \mathbf{y})$, $\mu(\boldsymbol{\theta})$ as follows

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

where \mathbf{b} is an n -dimensional vector with entries $\{b_i\}$ and \mathbf{D} is a diagonal matrix with n entries $\{c_i\}$. Note that both \mathbf{b} and \mathbf{D} depend on $\boldsymbol{\theta}$, so the Gaussian approximation is for a fixed $\boldsymbol{\theta}$.

Modern INLA



The process is iterated to find \boldsymbol{b} and \boldsymbol{D} that gives the Gaussian approximation at the mode, $\mu(\boldsymbol{\theta})$, so that

$$\boldsymbol{X}|\boldsymbol{\theta}, \boldsymbol{y} \sim N\left(\boldsymbol{\mu}(\boldsymbol{\theta}), \boldsymbol{Q}_{\boldsymbol{X}}^{-1}(\boldsymbol{\theta})\right).$$

The graph of the Gaussian approximation consists of two components,

- ① \mathcal{G}_p : the graph obtained from the prior of the latent field through $\boldsymbol{Q}(\boldsymbol{\theta})$
- ② \mathcal{G}_d : the graph obtained from the data based on the non-zero entries of $\boldsymbol{A}^\top \boldsymbol{A}$

How can we use VB?



We apply this to the Gaussian approximation in the denominator.

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

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



How can we use VB?

We apply this to the Gaussian approximation in the denominator.

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

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

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



How can we use VB?

We apply this to the Gaussian approximation in the denominator.

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

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



How can we use VB?

We apply this to the Gaussian approximation in the denominator.

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

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

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

Implicit mean correction



Recall that $Q\mu = b$.

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

$$\mu^* = \mu + M\lambda$$

Implicit mean correction



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

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

$$\mu^* = \mu + M\lambda$$

So now we solve for,

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

Implicit mean correction



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

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

$$\mu^* = \mu + M\lambda$$

So now we solve for,

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

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

Implicit mean correction



Recall that $Q\mu = b$.

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

$$\mu^* = \mu + M\lambda$$

So now we solve for,

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

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

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

Modern INLA



Next, the marginal conditional posteriors of the elements of \boldsymbol{X} is calculated from the joint Gaussian approximation as

$$\boldsymbol{X}_j | \boldsymbol{\theta}, \boldsymbol{y} \sim N \left((\boldsymbol{\mu}(\boldsymbol{\theta}))_j, (\boldsymbol{Q}_{\boldsymbol{X}}^{-1}(\boldsymbol{\theta}))_{jj} \right).$$

and the marginals

$$\tilde{\pi}(\boldsymbol{X}_j | \boldsymbol{y}) = \int \pi_G(\boldsymbol{X}_j | \boldsymbol{\theta}, \boldsymbol{y}) \tilde{\pi}(\boldsymbol{\theta} | \boldsymbol{y}) d\boldsymbol{\theta} \approx \sum_{k=1}^K \pi_G(\boldsymbol{X}_j | \boldsymbol{\theta}_k, \boldsymbol{y}) \tilde{\pi}(\boldsymbol{\theta}_k | \boldsymbol{y}) \delta_k.$$

Examples of LGMs



- Linear models
- Generalized linear models
- Generalized additive mixed models
 - Longitudinal data models
 - Overdispersion models
 - Time series models
 - Areal data models
 - State space models
 - Gaussian processes

Website for R-INLA library



Ethiopian Public Health Institute
ኢትዮጵያ የክብረተሰቦች ሚኒስቴር

<https://www.r-inla.org/>

Universal tools in INLA



Ethiopian Public Health Institute
ኢትዮጵያ የክርክሩ ሚኒስቴር

- 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

Latent field priors



Ethiopian Public Health Institute
ኢትዮጵያ ሬክተና አገልግሎት ቤት

- Gaussian with various Q

Hyperparameter priors



User's choice.

Default - mostly PC priors (Simpson et al. (2017)²)

²Simpson, D., Rue, H., Riebler, A., Martins, T.G. and Sørbye, S.H., 2017. Penalising model component complexity: A principled, practical approach to constructing priors.

Penalizing complexity priors



What is the purpose of the parameter? Function versus value.

- ① Occam's razor
- ② Measure of complexity KLD
- ③ Constant rate penalization
- ④ User-defined contraction scaling

KLD → distance → prior



Ethiopian Public Health Institute
ኢትዮጵያ ሬክሳይንስ በኋና አገልግሎት

$$\begin{aligned} \text{KLD}(\pi(\mathbf{x}|\xi) \| \pi(\mathbf{x}|\xi = 0)) \\ &= \int \pi(\mathbf{x}|\xi) \log \left(\frac{\pi(\mathbf{x}|\xi)}{\pi(\mathbf{x}|\xi = 0)} \right) d\mathbf{x}. \\ \pi(\xi) &= \lambda e^{-\lambda d(\xi)} \left| \frac{\partial d(\xi)}{\partial \xi} \right| \end{aligned}$$

Survival models as LGM's



Censoring and Truncation affects the likelihood only. For censoring,

- ① Right : $L_i(t_i|d_i) = S_i(t_i)$
- ② Event : $L_i(t_i|d_i) = S_i(t_i)f_i(t_i)$
- ③ Left : $L_i(t_i|d_i) = 1 - S_i(t_i)$
- ④ Interval : $L_i(t_{1i}, t_{2i}|d_i) = S_i(t_{1i}) - S_i(t_{2i}),$

Survival models as LGM's



Censoring and Truncation affects the likelihood only. For censoring,

- ① Right : $L_i(t_i|d_i) = S_i(t_i)$
- ② Event : $L_i(t_i|d_i) = S_i(t_i)f_i(t_i)$
- ③ Left : $L_i(t_i|d_i) = 1 - S_i(t_i)$
- ④ Interval : $L_i(t_{1i}, t_{2i}|d_i) = S_i(t_{1i}) - S_i(t_{2i}),$

so then (still)

Survival models as LGM's



Censoring and Truncation affects the likelihood only. For censoring,

- ① Right : $L_i(t_i|d_i) = S_i(t_i)$
- ② Event : $L_i(t_i|d_i) = S_i(t_i)f_i(t_i)$
- ③ Left : $L_i(t_i|d_i) = 1 - S_i(t_i)$
- ④ Interval : $L_i(t_{1i}, t_{2i}|d_i) = S_i(t_{1i}) - S_i(t_{2i}),$

so then (still)

- ① Data t_i with some likelihood $L(\mathbf{X}, \boldsymbol{\theta}|\mathbf{y}) = \prod_{i=1}^n L(t_i|\eta_i = h(\mathbf{X}, \boldsymbol{\theta}))$

Survival models as LGM's



Censoring and Truncation affects the likelihood only. For censoring,

- ① Right : $L_i(t_i|d_i) = S_i(t_i)$
- ② Event : $L_i(t_i|d_i) = S_i(t_i)f_i(t_i)$
- ③ Left : $L_i(t_i|d_i) = 1 - S_i(t_i)$
- ④ Interval : $L_i(t_{1i}, t_{2i}|d_i) = S_i(t_{1i}) - S_i(t_{2i}),$

so then (still)

- ① Data t_i with some likelihood $L(\mathbf{X}, \boldsymbol{\theta}|\mathbf{y}) = \prod_{i=1}^n L(t_i|\eta_i = h(\mathbf{X}, \boldsymbol{\theta}))$
- ② $\eta_i = h(\beta\mathbf{Z} + \mathbf{u}(\boldsymbol{\theta})\mathbf{A})$
- ③ $\mathbf{X}|\boldsymbol{\theta} \sim N(\mathbf{0}, \mathbf{Q}(\boldsymbol{\theta})^{-1})$
- ④ $\boldsymbol{\theta} \sim \text{hyperprior}$

Possibilities



- Frailty models - unit-specific or clustered
- Spatial survival models - areal or continuous in space
- Nonlinear effects of covariates using splines
- Joint models with continuous or discrete longitudinal biomarker(s)
- Competing risks or multi-state models
- and many more...

Acute Myeloid Leukemia study



Introduction

In this example we are studying the spatial distribution of leukemia mortality to inform public health policies, to gain insights for unmeasured covariates³



³van Niekerk, J. and Rue, H., 2024. Low-rank variational Bayes correction to the Laplace

Model



In this example we are studying the spatial distribution of leukemia mortality to inform public health policies, to gain insights for unmeasured covariates.

$$\begin{aligned} h(t|\beta, \boldsymbol{u}) &= h_0(t) \exp(\eta(s)) \\ \eta_i(s) &= \beta_0 + \beta_1 \text{Age}_i + \beta_2 \text{WBC}_i + \beta_3 \text{TPI}_i + u(s), \end{aligned}$$

to account for spatial variation we use a Gaussian effect \boldsymbol{u} with a Matérn covariance structure with hyperparameters, marginal variance σ_u^2 and nominal range $r = 2/\kappa$.



Posterior inference - fixed effects

Time used:

Pre = 4.1, Running = 2.58, Post = 0.156, Total = 6.84

Fixed effects:

	mean	sd	0.025quant	0.5quant	0.975quant	mode	kld
(Intercept)	-2.170	0.206	-2.569	-2.173	-1.753	-2.172	0
sex	0.072	0.069	-0.063	0.072	0.208	0.072	0
age	0.033	0.002	0.029	0.033	0.037	0.033	0
wbc	0.003	0.000	0.002	0.003	0.004	0.003	0
tpi	0.025	0.010	0.005	0.025	0.044	0.025	0

Random effects:

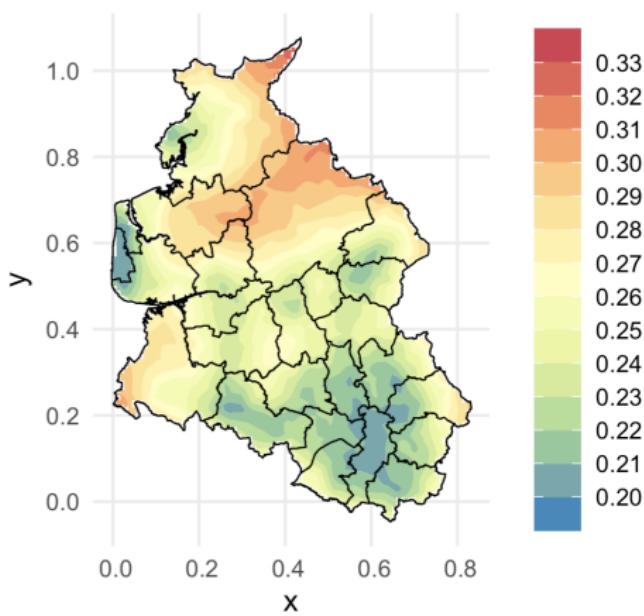
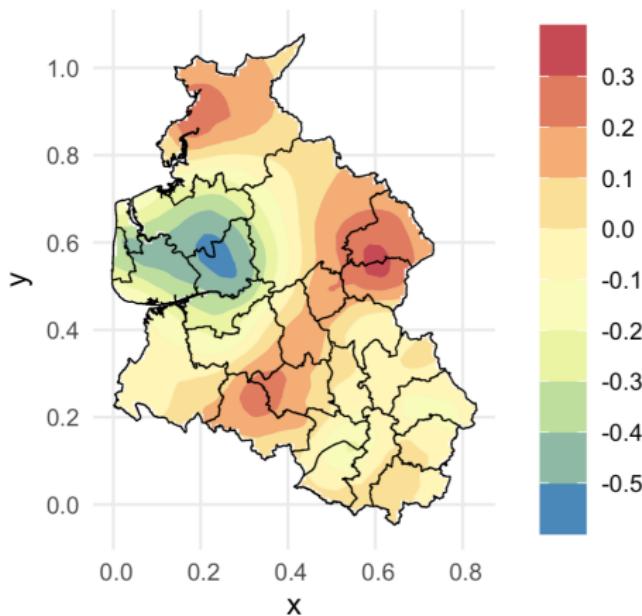
Name	Model
spatial	SPDE2 model

Model hyperparameters:

	mean	sd	0.025quant	0.5quant	0.975quant	mode
alpha parameter for weibullsurv	0.599	0.016	0.568	0.599	0.631	0.599
Range for spatial	0.310	0.156	0.114	0.276	0.709	0.220
Stdev for spatial	0.293	0.073	0.174	0.284	0.460	0.268

Marginal log-Likelihood: -839.92

Posterior inference - spatial field



Posterior inference - with MCMC



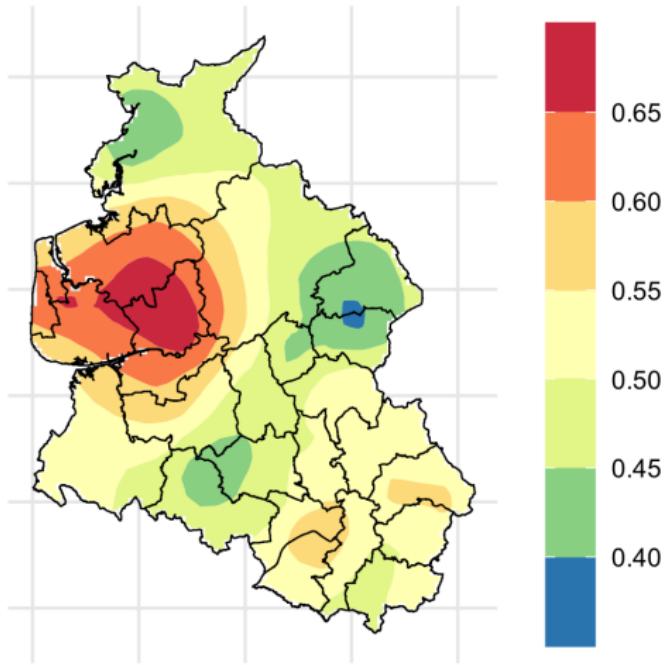
	HMC	INLA
β_0	-2.189	-2.189
β_1	0.597	0.597
β_2	0.241	0.241
β_3	0.108	0.108
τ	0.340	0.340
σ_u	0.223	0.223
r	0.202	0.202
Time(s)	8214	6.8

Table: Posterior means from HMC and INLA

Survival probabilities



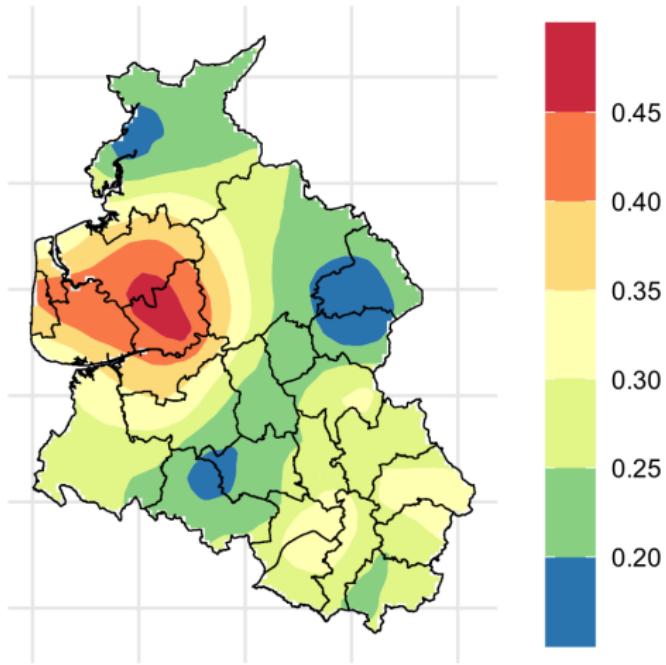
Survival function at year 1



Survival probabilities



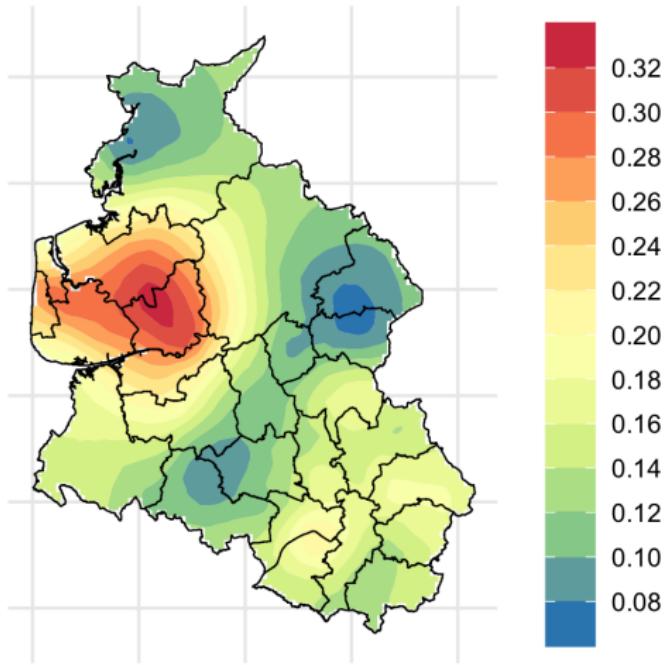
Survival function at year 2



Survival probabilities



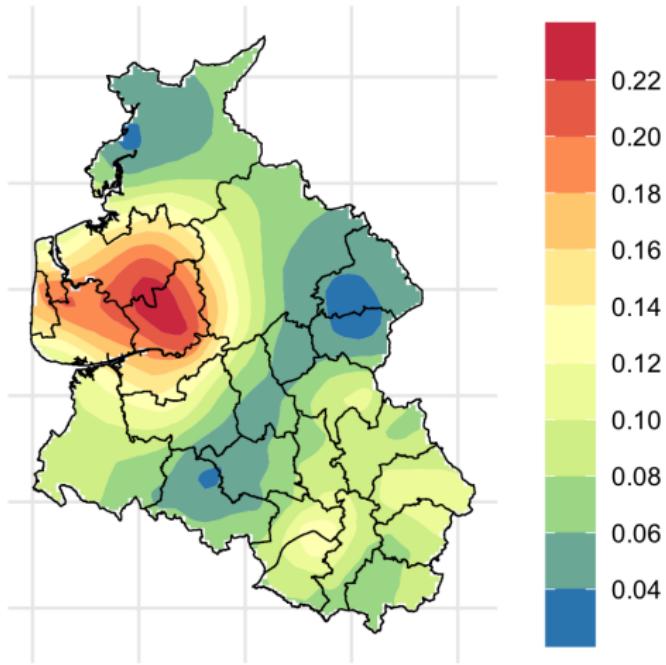
Survival function at year 3



Survival probabilities



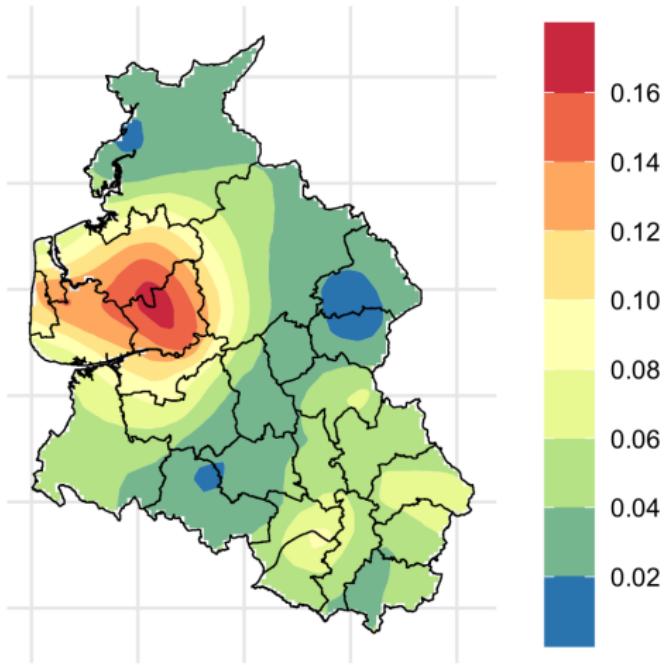
Survival function at year 4



Survival probabilities



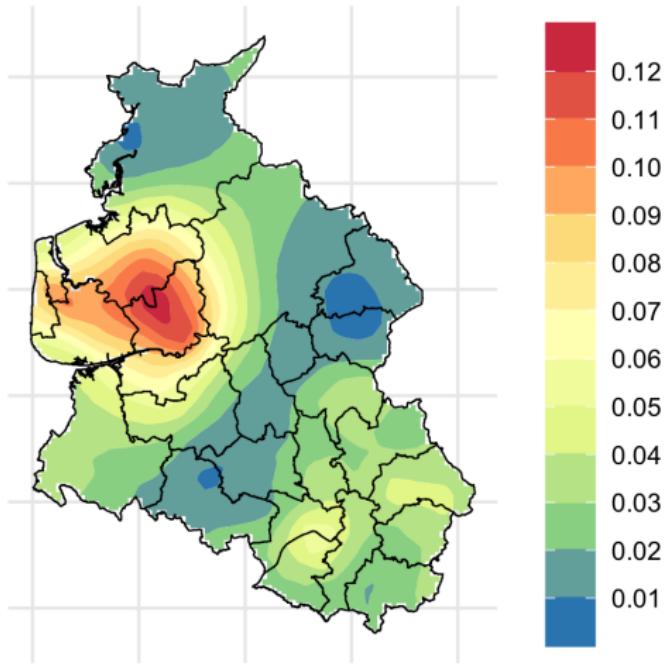
Survival function at year 5



Survival probabilities



Survival function at year 6



Medication adherence study

Introduction



From the AARDEX Group - medication adherence and patient persistence is crucial for successful treatment regimes and regularized evaluation⁴. Adherence is a proportion and persistence is defined based on a set of criteria unique to each drug.

⁴Burger, D.A., Van der Merwe, S., Van Niekerk, J., Lesaffre, E. and Pironet, A. Joint quantile regression of longitudinal continuous proportions and time-to-event data: application in medication adherence and persistence, *Statistical Methods in Medical Research*, Accepted

Introduction



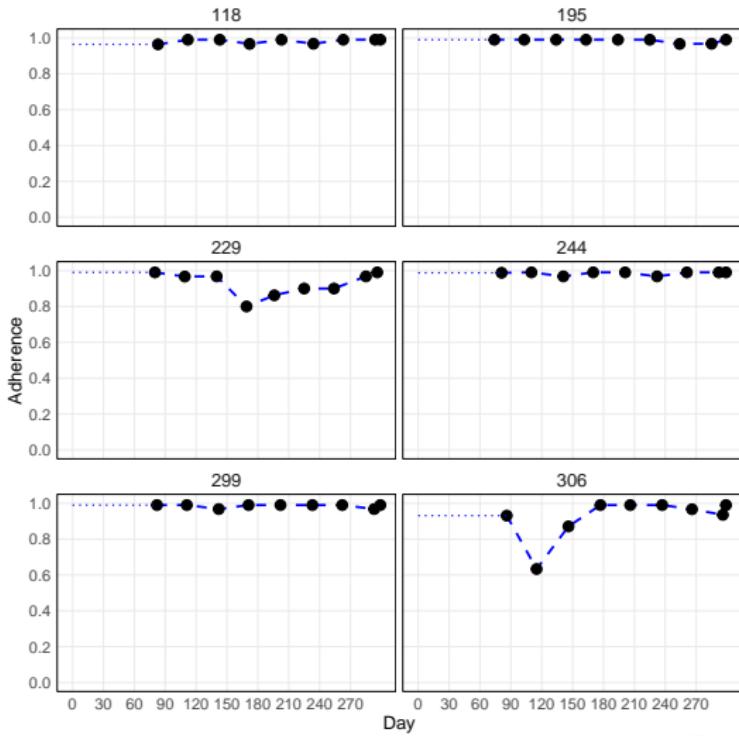
From the AARDEX Group - medication adherence and patient persistence is crucial for successful treatment regimes and regularized evaluation⁴.

Adherence is a proportion and persistence is defined based on a set of criteria unique to each drug.

Research problem - outliers in the adherence bias the model, and can cause false positives for intervention.

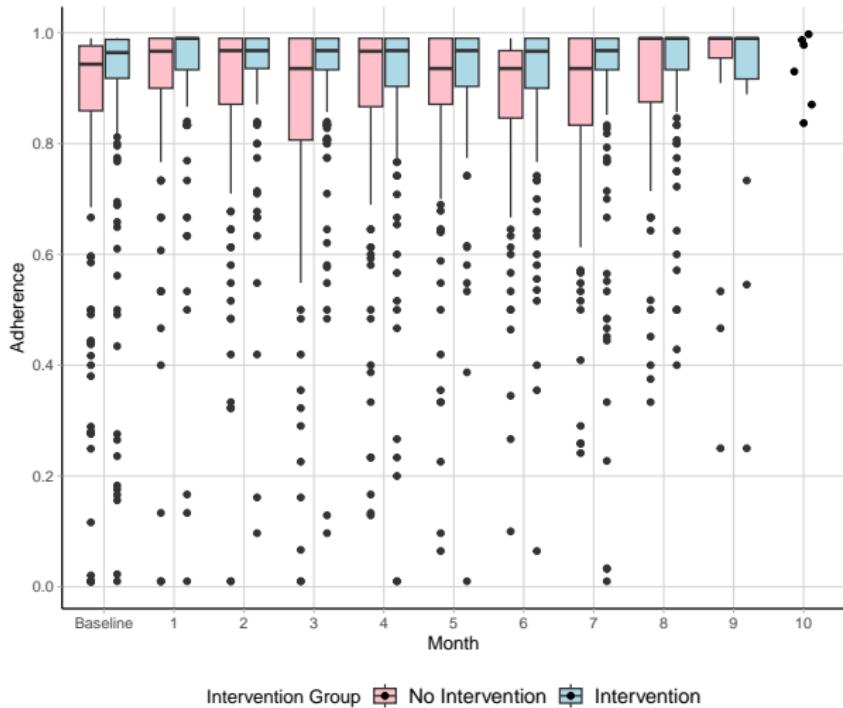
⁴Burger, D.A., Van der Merwe, S., Van Niekerk, J., Lesaffre, E. and Pironet, A. Joint quantile regression of longitudinal continuous proportions and time-to-event data: application in medication adherence and persistence, *Statistical Methods in Medical Research*, Accepted

Dataset





Dataset



Model



Longitudinal model:

$$f(y_{ij}; \kappa_{q,i}(t_{ij}), \psi) = \alpha_{ij_1} \alpha_2 y_{ij}^{\alpha_{ij_1}-1} \left(1 - y_{ij}^{\alpha_{ij_1}}\right)^{\alpha_2-1},$$

Model



Longitudinal model:

$$\begin{aligned}f(y_{ij}; \kappa_{q,i}(t_{ij}), \psi) &= \alpha_{ij_1} \alpha_2 y_{ij}^{\alpha_{ij_1}-1} \left(1 - y_{ij}^{\alpha_{ij_1}}\right)^{\alpha_2-1}, \\ \kappa_{q,i}(t_{ij}) &= g^{-1}(\eta_{q,i}(t_{ij}))\end{aligned}$$

Model



Longitudinal model:

$$f(y_{ij}; \kappa_{q,i}(t_{ij}), \psi) = \alpha_{ij_1} \alpha_2 y_{ij}^{\alpha_{ij_1}-1} \left(1 - y_{ij}^{\alpha_{ij_1}}\right)^{\alpha_2-1},$$

$$\kappa_{q,i}(t_{ij}) = g^{-1}(\eta_{q,i}(t_{ij}))$$

$$\eta_{q,i}(t_{ij}) = (\beta_0 + b_{0_i}) + (\beta_{\text{time}} + b_{\text{time}_i} + \mathbf{z}'_i \boldsymbol{\beta}_{\text{tx}}) t_{ij} + \mathbf{x}'_i \boldsymbol{\beta}_{\text{cov}}$$

with $\alpha_{ij_1} = \frac{\log\left(1-(1-q)^{\frac{1}{\alpha_2}}\right)}{\log(\kappa_{q,i}(t_{ij}))}$ and $\alpha_2 = \frac{\log(1-q)}{\log(1-e^{-\psi})}$.

Model



Longitudinal model:

$$f(y_{ij}; \kappa_{q,i}(t_{ij}), \psi) = \alpha_{ij_1} \alpha_2 y_{ij}^{\alpha_{ij_1}-1} \left(1 - y_{ij}^{\alpha_{ij_1}}\right)^{\alpha_2-1},$$

$$\kappa_{q,i}(t_{ij}) = g^{-1}(\eta_{q,i}(t_{ij}))$$

$$\eta_{q,i}(t_{ij}) = (\beta_0 + b_{0i}) + (\beta_{\text{time}} + b_{\text{time}_i} + \mathbf{z}'_i \boldsymbol{\beta}_{\text{tx}}) t_{ij} + \mathbf{x}'_i \boldsymbol{\beta}_{\text{cov}}$$

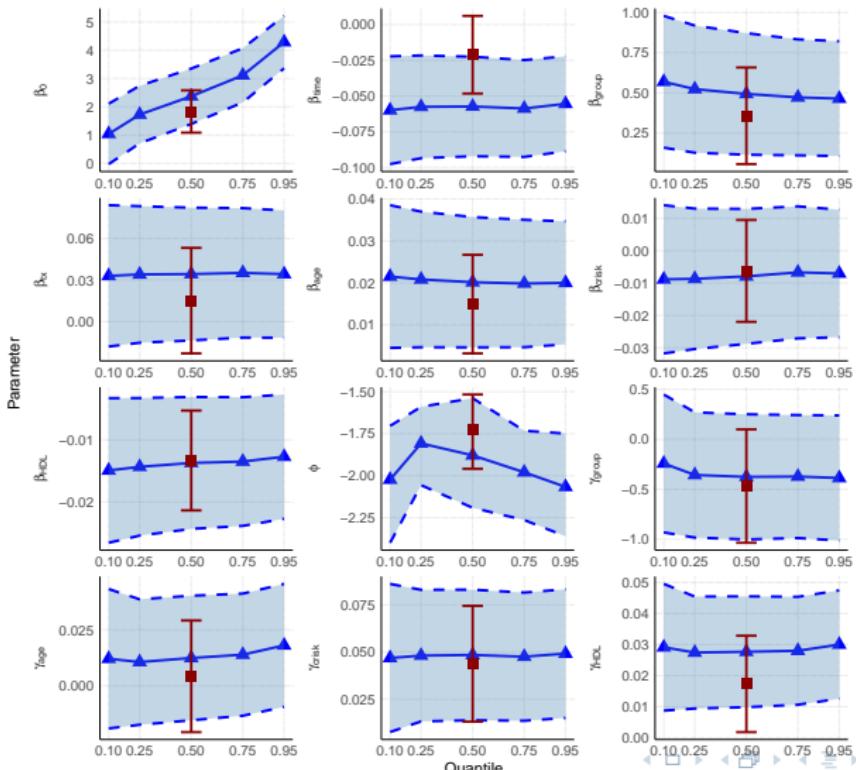
with $\alpha_{ij_1} = \frac{\log\left(1-(1-q)^{\frac{1}{\alpha_2}}\right)}{\log(\kappa_{q,i}(t_{ij}))}$ and $\alpha_2 = \frac{\log(1-q)}{\log(1-e^{-\psi})}$.

Survival model:

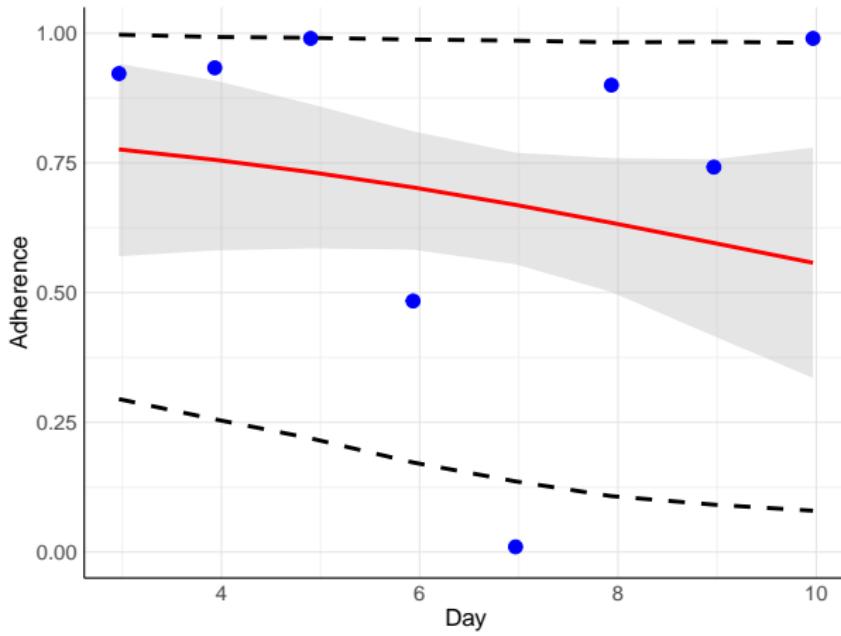
$$h_i(t) = h_0(t) \exp(\phi \eta_i(t) + \mathbf{w}'_i \boldsymbol{\gamma}).$$



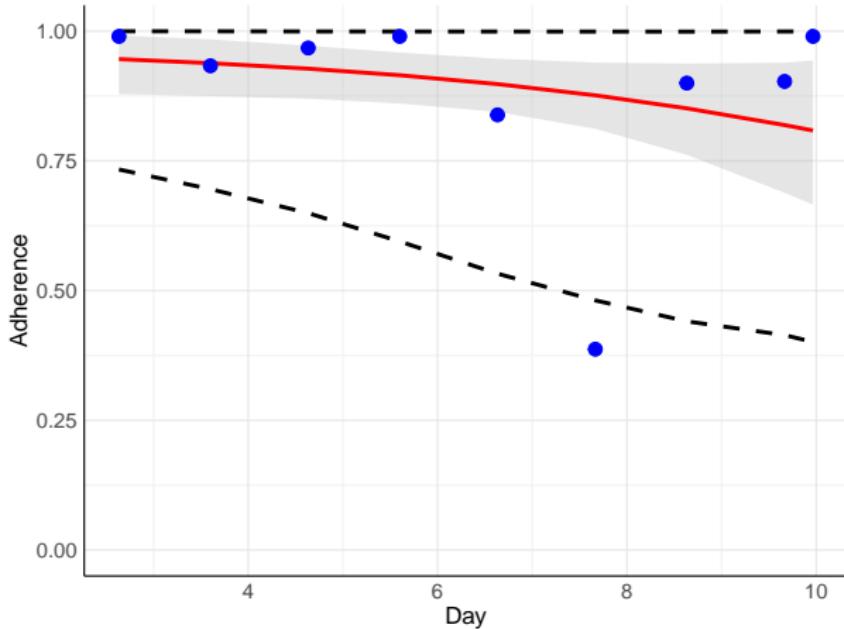
Posterior inference - fixed effects



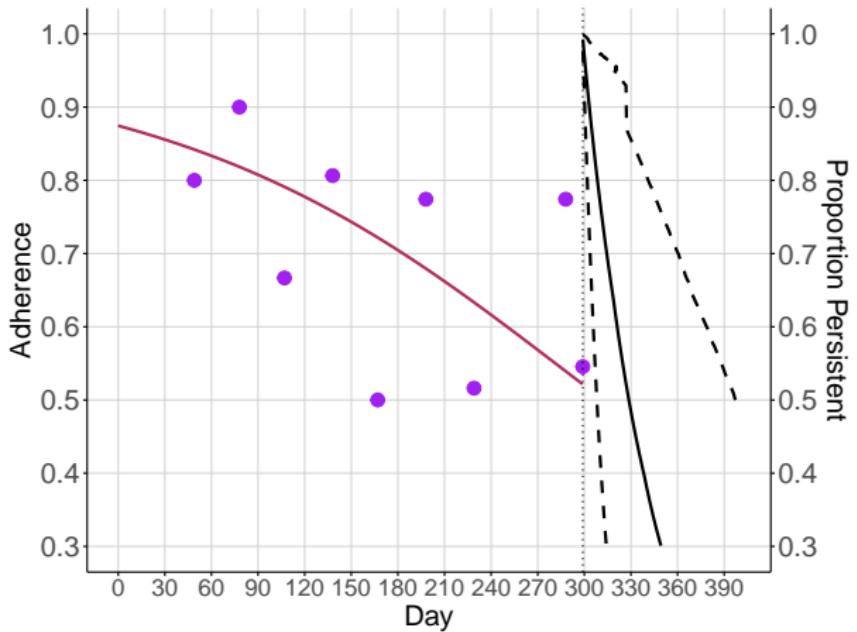
Posterior inference - Patient 46



Posterior inference - Patient 56



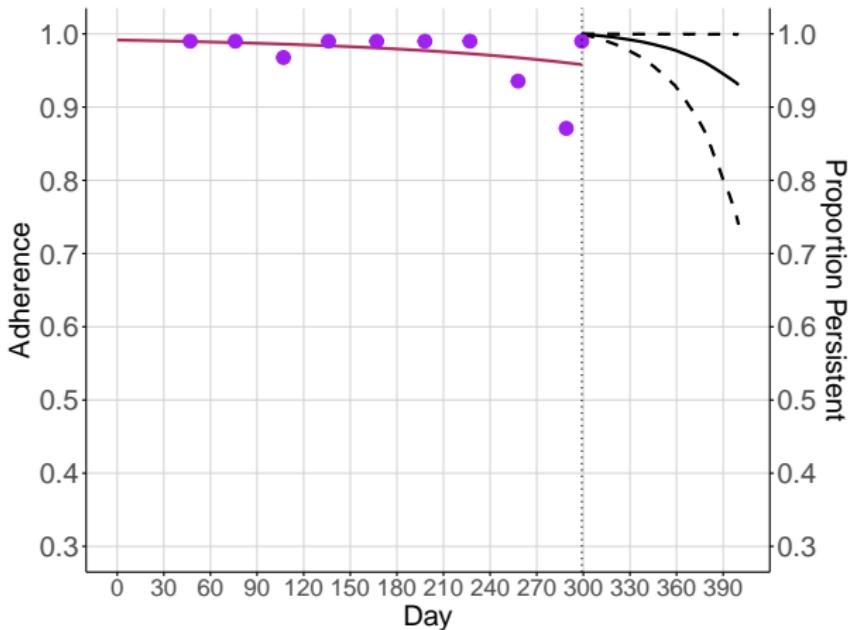
Dynamic predictions - Patient 27



Dynamic predictions - Patient 10



Ethiopian Public Health Institute
ኢትዮጵያ ሬክተና አገልግሎት



Dengue risk in Brazil

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,

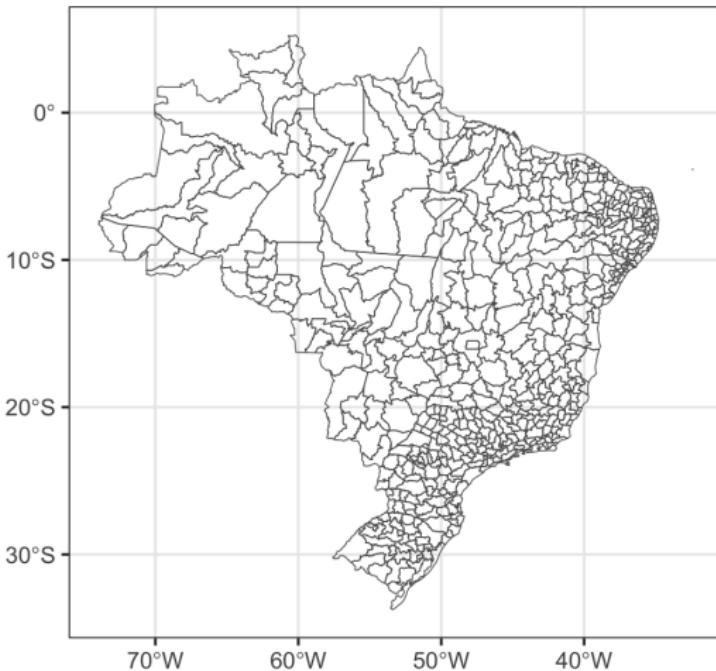
$$\mathbf{y} \sim \text{Poisson}(Ee^{\boldsymbol{\eta}}), \quad \boldsymbol{\eta} = \mathbf{1}^T \boldsymbol{\mu} + \boldsymbol{\alpha}$$

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. We have 561000+ cases for a year.

Dengue risk in Brazil



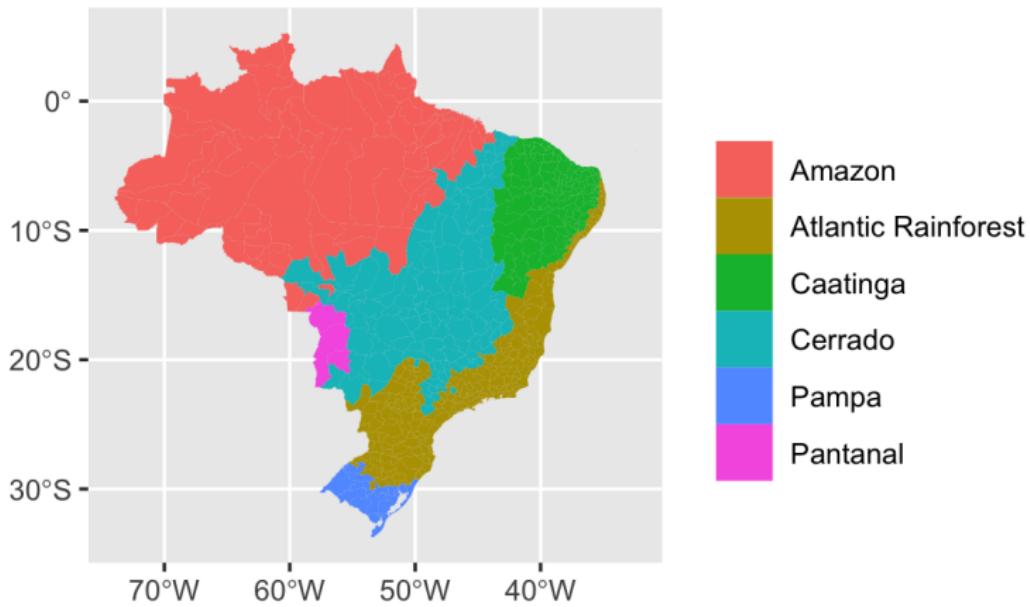
Ethiopian Public Health Institute
ኢትዮጵያ ሬክተና አገልግሎት
Ethiopian Public Health Institute



Dengue risk in Brazil



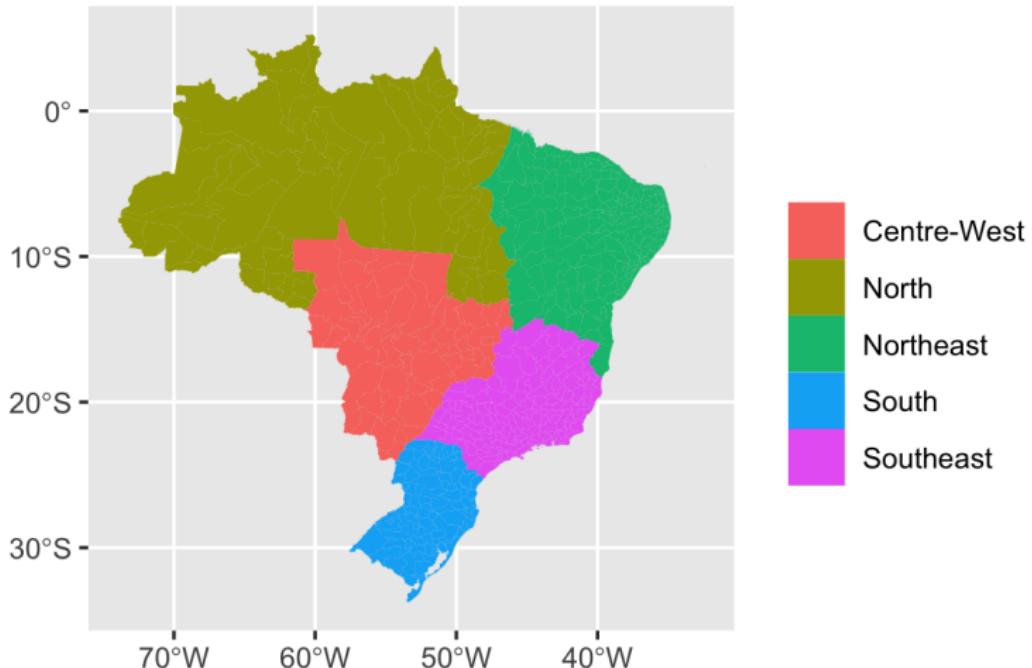
Ethiopian Public Health Institute
ኢትዮጵያ ሬክተና አገልግሎት
Ethiopian Public Health Institute



Dengue risk in Brazil



Ethiopian Public Health Institute
ኢትዮጵያ ሬክተና አገልግሎት
Ethiopian Public Health Institute



Flexible Besag model⁵



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

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

⁵Abdul-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

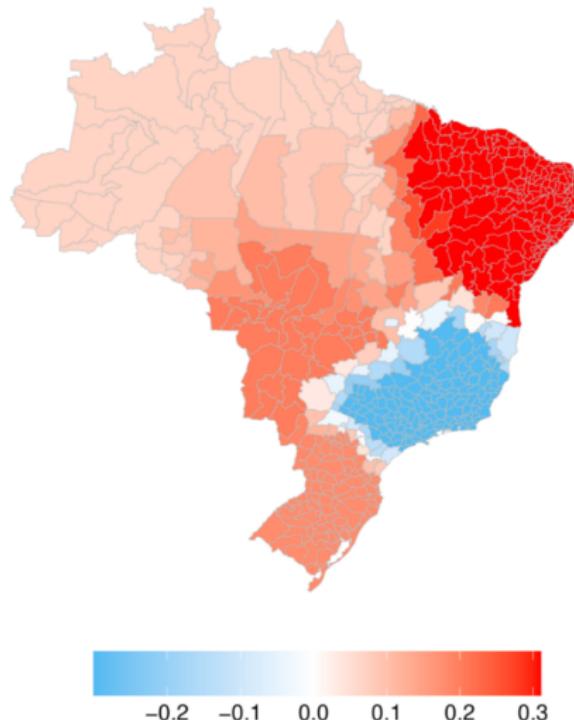
The joint PC prior for $\boldsymbol{\theta} = \log \boldsymbol{\tau}$ can be derived as a convolution of the PC prior for $\boldsymbol{\tau}$ from the Besag model, as follows

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

This prior contracts

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

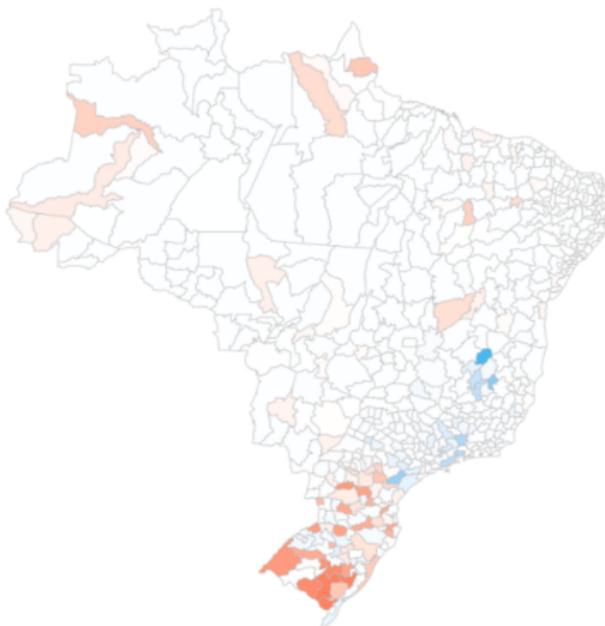
Results

(a) $\log \tau$

Results



Ethiopian Public Health Institute
ኢትዮጵያ ሬክተና አገልግሎት
Ethiopian Public Health Institute

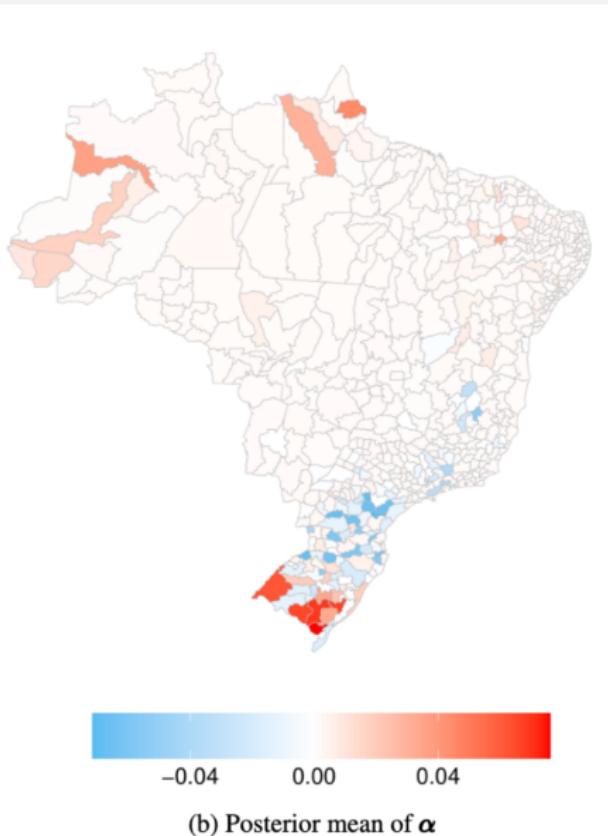


(b) Posterior mean of α

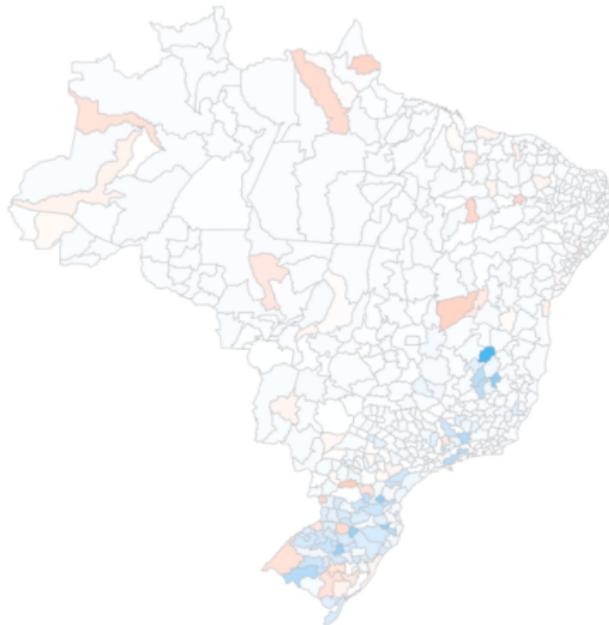
Difference in results



Ethiopian Public Health Institute
ኢትዮጵያ ሬክተና አገልግሎት በኋላ አገልግሎት

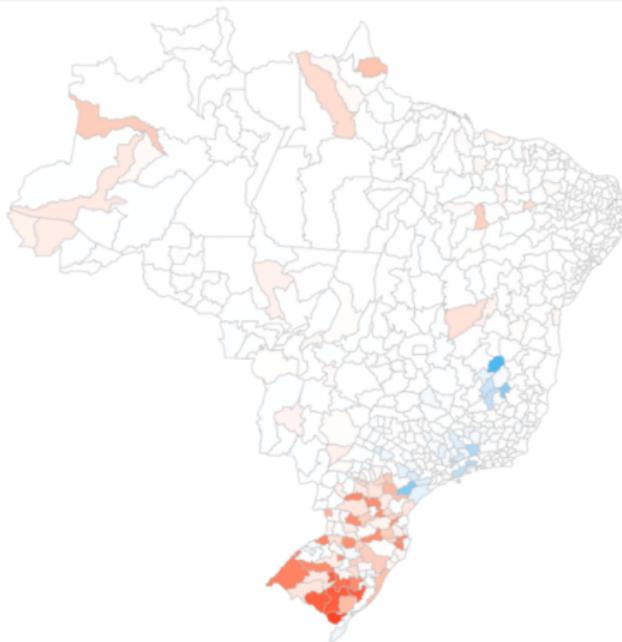


Difference in results



(c) 97.5th posterior percentile of α

Difference in results

(d) 2.5th posterior percentile of α

Dementia study



cs-fMRI model

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:

$$\mathbf{y}|\boldsymbol{\beta}, \mathbf{b}, \boldsymbol{\theta} \sim N(\boldsymbol{\mu}_y, \mathbf{V}), \quad \boldsymbol{\mu}_y = \sum_{k=0}^K \mathbf{X}_k \boldsymbol{\beta}_k + \sum_{j=1}^J \mathbf{Z}_j \mathbf{b}_j$$

$$\boldsymbol{\beta}_k = \boldsymbol{\Psi}_k \mathbf{w}_k \quad (\text{SPDE prior on } \boldsymbol{\beta}_k)$$

$$\mathbf{w}_k | \boldsymbol{\theta} \sim N(\mathbf{0}, \mathbf{Q}_{\tau_k, \kappa_k}^{-1})$$

$$\mathbf{b}_j \sim N(\mathbf{0}, \delta I) \quad (\text{Diffuse priors for } \mathbf{b}_j)$$

$$\boldsymbol{\theta} \sim \pi(\boldsymbol{\theta}),$$

where we have K task signals and J nuisance signals.⁶

⁶Van Niekerk, J., Krainski, E., Rustand, D. and Rue, H., 2023. A new avenue for Bayesian inference with INLA. Computational Statistics & Data Analysis, 181, p.107692.

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 y of size **2 523 624**, with an SPDE model defined on a mesh with 8795 triangles⁷.

⁷Mejia, A.F., Yue, Y., Bolin, D., Lindgren, F. and Lindquist, M.A., 2020. A Bayesian general linear modeling approach to cortical surface fMRI data analysis. *Journal of the American Statistical Association*, 115(530), pp.501-520.

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 y of size **2 523 624**, with an SPDE model defined on a mesh with 8795 triangles⁷.

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

⁷ Mejia, A.F., Yue, Y., Bolin, D., Lindgren, F. and Lindquist, M.A., 2020. A Bayesian general linear modeling approach to cortical surface fMRI data analysis. *Journal of the American Statistical Association*, 115(530), pp.501-520.

cs-fMRI model

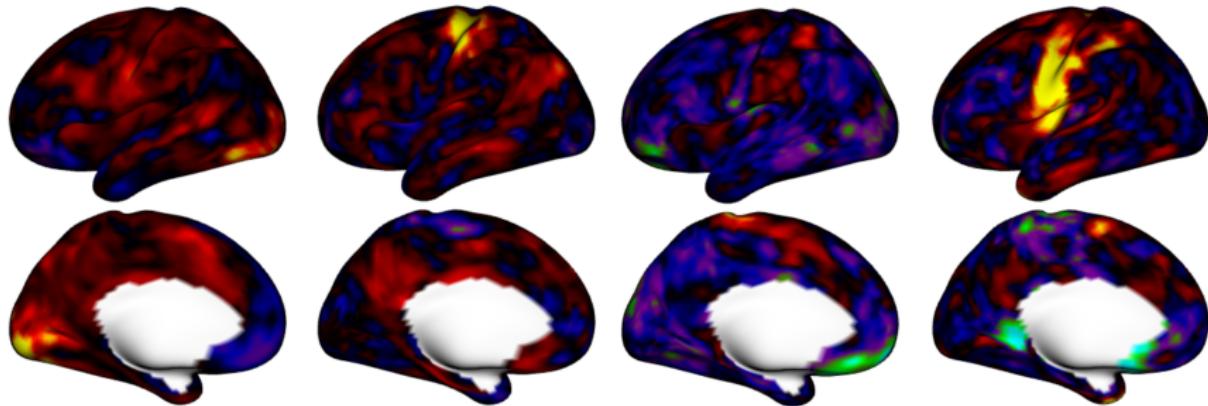


Figure: Activation areas for the different tasks in the left hemisphere - visual cue, right hand motor, right foot motor, tongue motor task (from left to right)



Further reading

- 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), pp.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, p.107692.
- 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.
- 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), p.25.



شكراً • Thank you



جامعة الملك عبد الله
للعلوم والتكنولوجيا
King Abdullah University of
Science and Technology