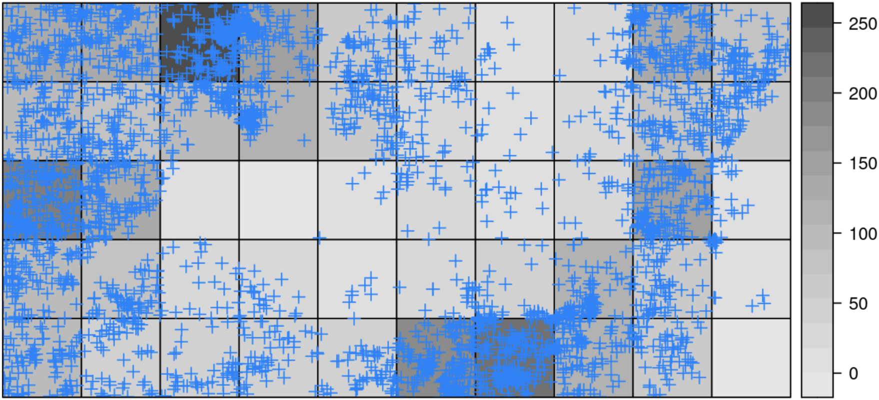
Spatial modeling with INLA

King Abdullah University of Science and Technology
جامعة الملك عبدالله
للعلوم والتقنية

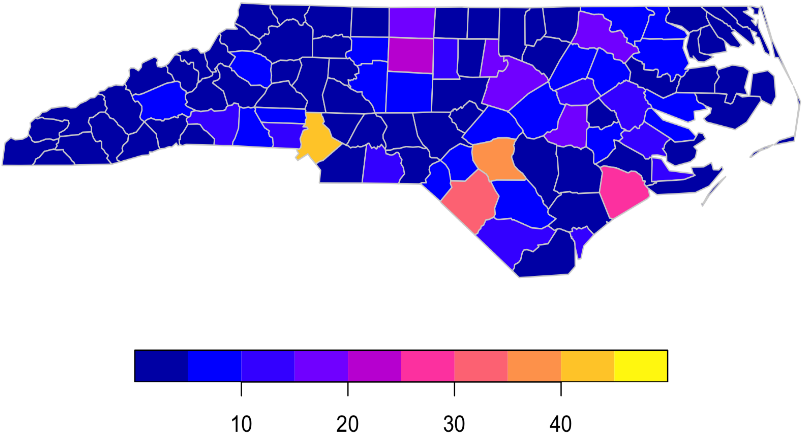
July 2024
The repo for this workshop is here
https://tinyurl.com/INLAUP24

Outline I

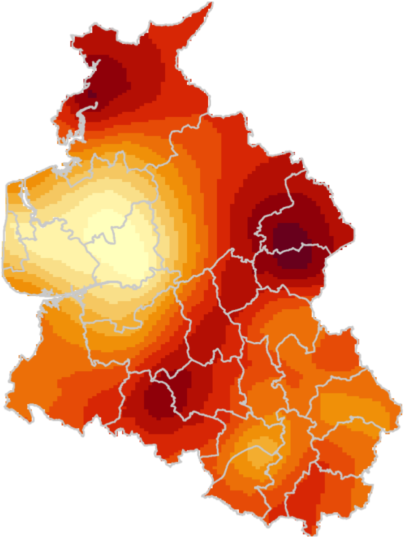
Lattice type - Day 2



Irregular lattice - areal data - Day 2



Continuous domain - geostatistics - Day 2

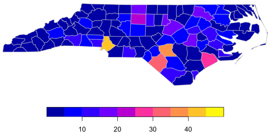


Point process - Day 3

Universal tools in INLA - also for spatial models

- 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

Besag and BYM



Besag model is a "smoother" over space.

$$x_i | \boldsymbol{x}_{-i} \sim N\left(\frac{1}{n_i}\sum_{i\sim j} x_j, \frac{1}{n_i\tau}\right)$$

BYM (Besag + iid) parameterized for intrepretable parameters.

$$x_i = \frac{1}{\sqrt{\tau}}\left(\sqrt{\phi}u_i + \sqrt{1-\phi}v_i\right).$$

North Carolina SIDS example

NC sids R markdown file - Example 1 and 2

Malaria and G6PD example on joint spatial modeling

We can do joint modeling and quantile models as well.[1]

Malaria and G6PD R markdown file - Example 3 and 4

[1]Alahmadi, H., Van Niekerk, J., Padellini, T. and Rue, H., 2024. Joint quantile disease mapping with application to malaria and G6PD deficiency. Royal Society Open Science, 11(1), p.230851.

Flexible Besag model[2]

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

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

and

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

[2] 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 $\rightarrow$ stationary

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

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

This prior contracts

$$\tau_1, \tau_2, \ldots, \tau_P \quad \rightarrow \quad \tau$$

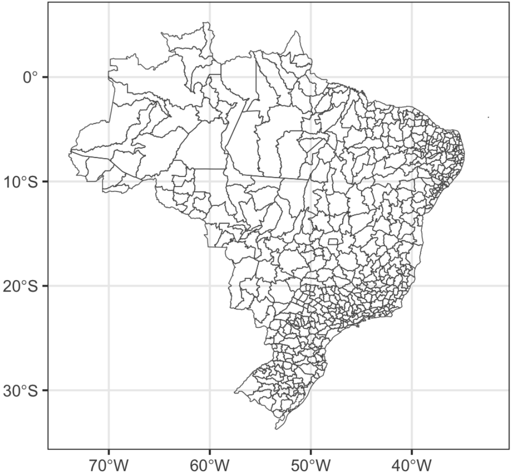
Dengue risk in Brazil

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,

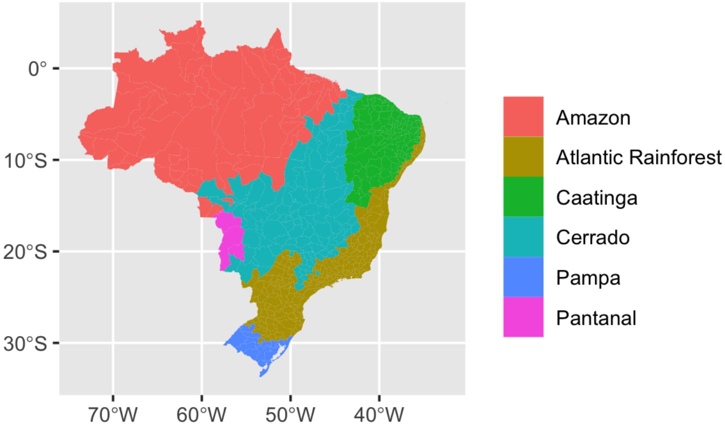
$$\boldsymbol{y} \sim \text{Poisson}(E e^{\boldsymbol{\eta}}), \quad \boldsymbol{\eta} = \mathbf{1}^T \mu + \boldsymbol{\alpha}$$

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

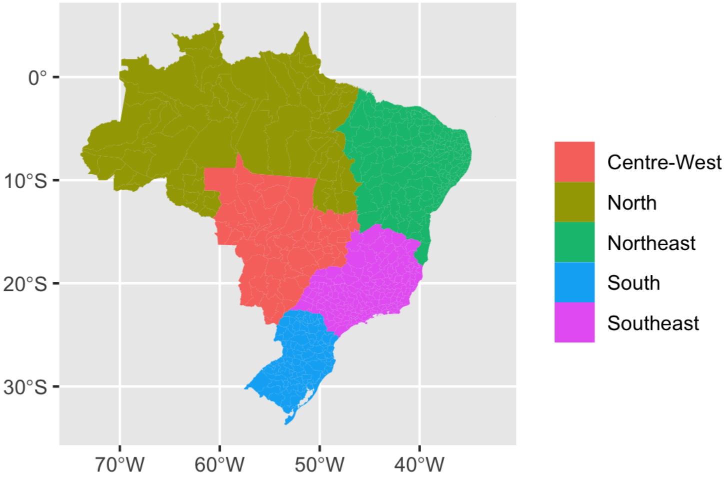
Dengue risk in Brazil



Dengue risk in Brazil



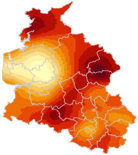
Dengue risk in Brazil



Dengue risk in Brazil

Dengue risk in Brazil R markdown file - Example 5

Each of these spatial models can be used to form a space-time interaction.
In the areal modeling framework this is done with the Kronecker product of the precision matrices of the spatial and temporal components.
NOTE: These models need a lot of constraints to ensure identifiablity.
Details for current models are here: Knorr-Held models



Kriging provides conditional expectations of the spatial field based on covariance parameters.
With INLA we estimate "covariance" parameters in a Bayesian way and provide the marginal expectation of the spatial field.
So INLA also does "Kriging" and more - Kriging is a model, not a method.

Matérn field

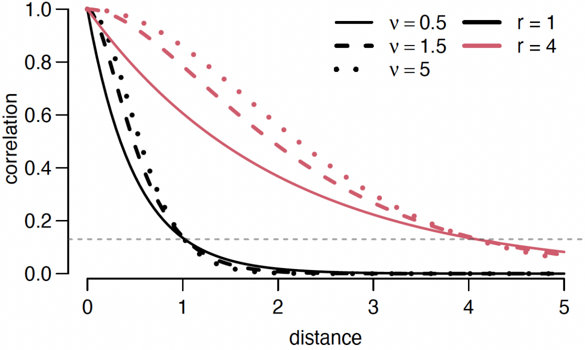
Consider a set of locations $\boldsymbol{s}$, then the spatial field $\boldsymbol{u}$ defined at $\boldsymbol{s}$ is multivariate Gaussian with the Matérn covariance function for the elements of $\Sigma(\theta)$,

$$\pi(\mathbf{u}|\theta) = (2\pi)^{-n/2}|\Sigma(\theta)|^{-1/2} \exp\left(-\frac{1}{2}\mathbf{u}^T\Sigma(\theta)^{-1}\mathbf{u}\right)$$

Matérn covariance model

$$\text{Matérn(1960)}: \ \Sigma_{ij} = \frac{\sigma^2(\kappa\|\mathbf{s}_i - \mathbf{s}_j\|)^\nu K_\nu(\kappa\|\mathbf{s}_i - \mathbf{s}_j\|)}{\Gamma(\nu + d/2)(4\pi)^{d/2}\kappa^{2\nu}2^{\nu-1}}$$

If $d = 2$ and $\nu = 1$: Whittle (1954)



practical range $= \mathrm{r} = \sqrt{8\nu}/\kappa$, $\mathrm{corr(r)} \approx 0.13$

The Matérn's SPDE

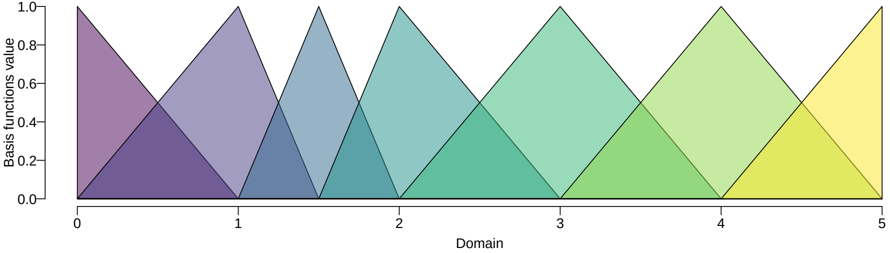
- Whittle (1954), Whittle (1963):
  - Fields with Matérn covariance are solutions to the following Stochastic Partial Differential Equation (SPDE)

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

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

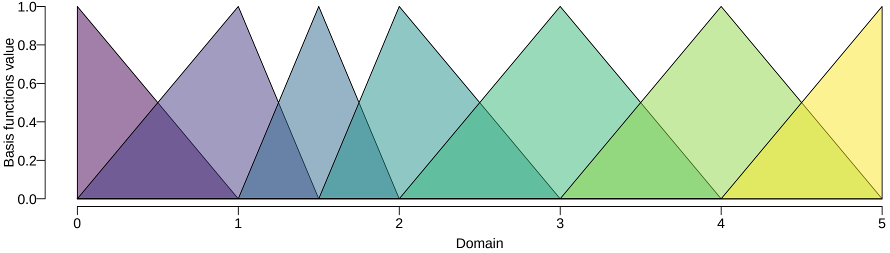
$$\Delta = \sum_{i=1}^{d} \frac{\partial^2}{\partial s_i^2}$$

How to solve the SPDE? FEM



- $u(\mathbf{s}) \approx \sum_{k=1}^{m} \psi_k(\mathbf{s}) u_k = \mathbf{A}(\mathbf{s}, \mathbf{s}_0) u(\mathbf{s}_0)$,
  - ψ_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



- $u(\mathbf{s}) \approx \sum_{k=1}^{m} \psi_k(\mathbf{s}) u_k = \mathbf{A}(\mathbf{s}, \mathbf{s}_0) u(\mathbf{s}_0)$,
    - ψ_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)[3] I

- Discretization
  - sparse precision matrix:
  - $\mathbf{Q}_{\alpha}(\tau, \kappa)$, for $\alpha \in \{1, 2, \ldots\}$.
- α
  - $\alpha = 1$: $\tau \mathbf{K}_1(\kappa) = \tau(\kappa^2 \mathbf{C} + \mathbf{G})$
  - $\alpha = 2$: $\tau(\kappa^4 \mathbf{C} + 2\kappa^2 \mathbf{G} + \mathbf{G}\mathbf{C}^{-1}\mathbf{G})$
  - $\alpha = 2, 3, 4, \ldots$: $\tau \mathbf{K}_1(\kappa)\mathbf{C}^{-1}\mathbf{K}_{\alpha-2}(\kappa)\mathbf{C}^{-1}\mathbf{K}_1(\kappa)$

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)[4] II

- d=1, $u_1, u_2, ..., u_n$, two neighbours

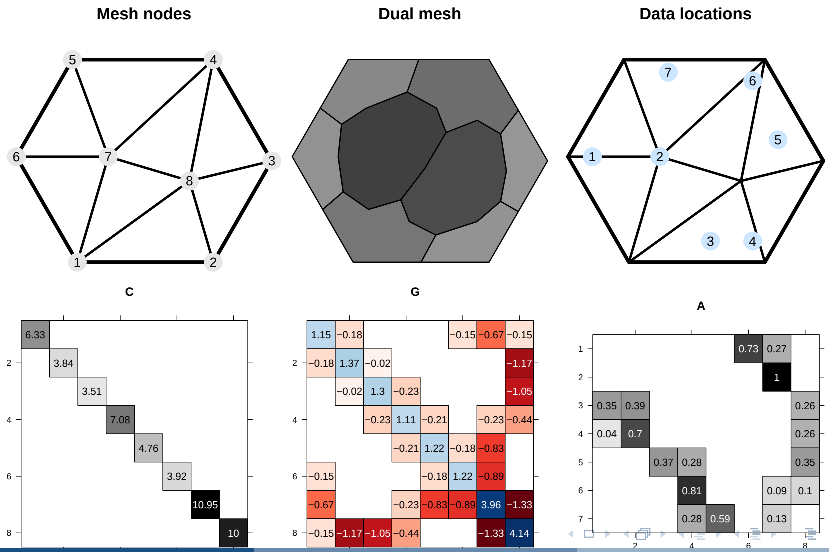
$$\tau^2 \begin{bmatrix} 1+\kappa^2 & -1 & & & \\ -1 & 2+\kappa^2 & -1 & & \\ & & \ddots & & \\ & & -1 & 2+\kappa^2 & -1 \\ & & & -1 & 1+\kappa^2 \end{bmatrix}$$

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

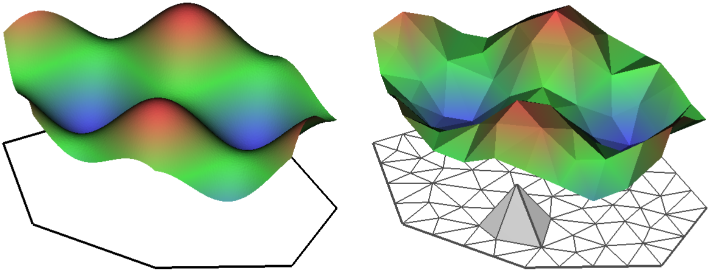
[3] Lindgren, F., Rue, H. and Lindström, J., 2011. An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. Journal of the Royal Statistical Society Series B: Statistical Methodology, 73(4), pp.423-498.

[4] 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



FEM in 3D



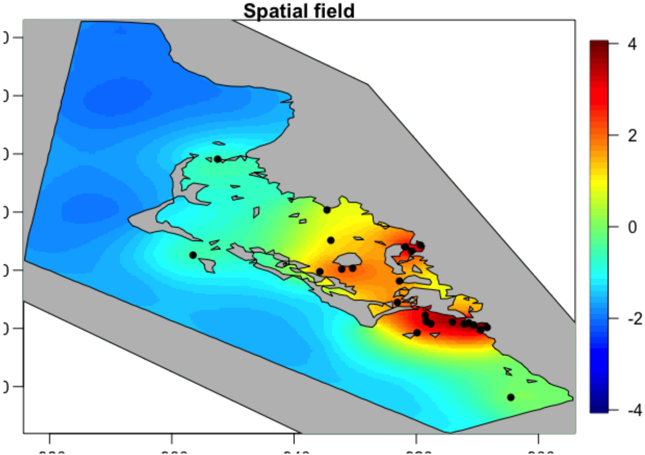
Geostatistical survival analysis

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

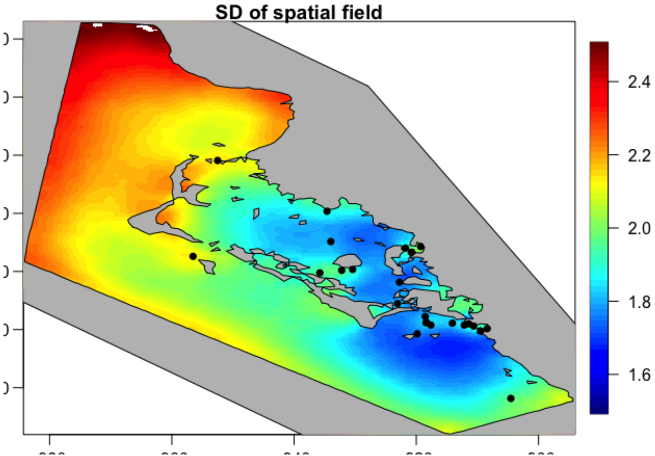
Leukemia mortality example in R - Example 1

Non-stationary Matern field based on physical barriers

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



Non-stationary Matern field based on physical barriers



Spatially varying coefficient models

We will use inlabru here -SVC using inlabru

$$\eta_{st} = \kappa_s + \alpha_s + \epsilon_s X_{1,st} + \tau_s X_{2,st} \tag{1}$$

Spatio-temporal models

Rainfall example in Parana state with a hurdle model. In space we have a Matern model and in time we use an AR(1) model. So this is a seperable model. For non-seperable models see here[5].

[5] Lindgren, F., Bakka, H., Bolin, D., Krainski, E. and Rue, H., 2024. A diffusion-based spatio-temporal extension of Gaussian Matérn fields. SORT-Statistics and Operations Research Transactions, pp.3-66.

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.

Thank you • شكرا

جامعة الملك عبدالله للعلوم والتقنية
King Abdullah University of Science and Technology