An introduction to INLA for spatial modeling

Joaquin Cavieres

Instituto de Estadística, Universidad de Valparaíso

Contents

Universidad de Valparaíso CHILE

Overview

What is the spatial statistics?

Let to consider a spatial process in $d = 2$ defined by:

$$\{Y(\boldsymbol{s}) : \boldsymbol{s} \in D \subset \mathbb{R}^d\}, \quad (1)$$

where Y is the observed variable, for example, the number of sick in a commune or in a neighborhood, or the rainfall in a region. We can denote to $\boldsymbol{s}$ as the geographical site were was measured that observation and D is a subset $\mathbb{R}$.

What is the spatial statistics?

Let to consider a spatial process in $d = 2$ defined by:

$$\{Y(\boldsymbol{s}) : \boldsymbol{s} \in D \subset \mathbb{R}^d\}, \quad (1)$$

where Y is the observed variable, for example, the number of sick in a commune or in a neighborhood, or the rainfall in a region. We can denote to $\boldsymbol{s}$ as the geographical site were was measured that observation and D is a subset $\mathbb{R}$.

Cressie (1993) propose three types of spatial observations:

1. **Areal data**

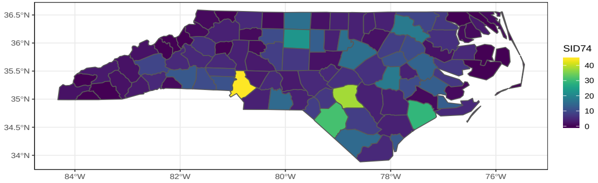
In the areal data (also known as *lattice data*) the spatial domain D is fixed and it's partitioned into a finite number of areal units with well-defined boundaries.

Cressie (1993) propose three types of spatial observations:

1. **Areal data**

In the areal data (also known as *lattice data*) the spatial domain D is fixed and it's partitioned into a finite number of areal units with well-defined boundaries.

Sudden infant deaths in North Carolina in 1974 (Pebesma (2018); Moraga (2019))



Cressie (1993) propose three types of spatial observations:

2. **Point patterns**

In this type of observations the spatial domain D is random. A set of indexes provides the locations of random events which are the pattern of the spatial points.

Cressie (1993) propose three types of spatial observations:

2. **Point patterns**

In this type of observations the spatial domain D is random. A set of indexes provides the locations of random events which are the pattern of the spatial points.

Universidad de Valparaíso CHILE

John Snows map of the 1854 London cholera outbreak (Moraga (2019))

Cressie (1993) propose three types of spatial observations:

3. **Geostatistical data**

In the geostatistical data the spatial domain D is continuous and fixed, this mean, $\boldsymbol{s}$ varies continually through D and $Y(\boldsymbol{s})$ can be observed in any place of D. The continuity is only for the domain and $Y(\boldsymbol{s})$ can be a continuous or discrete variable.

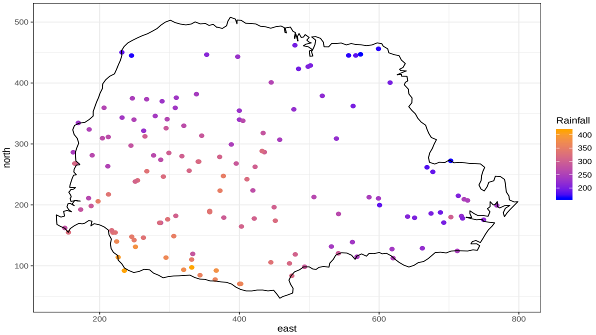
Cressie (1993) propose three types of spatial observations:

3. **Geostatistical data**

In the geostatistical data the spatial domain D is continuous and fixed, this mean, $\boldsymbol{s}$ varies continually through D and $Y(\boldsymbol{s})$ can be observed in any place of D. The continuity is only for the domain and $Y(\boldsymbol{s})$ can be a continuous or discrete variable.

Universidad de Valparaíso CHILE

Average rainfall measured at 143 recording stations in Paraná state, Brazil (Moraga (2019))



There are many `R` packages to model/predict observations spatially measured, for example:

- `sf`
- `geoR`
- `geoRglm`
- `GMRFLib`
- `RandomFields`
- `gstat`
- `rgdal`
- `GeoModels`
- ......
- `INLA`

In the package `INLA` you can build statistical models for the three spatial observations mentioned before, but here we will focus in the geostatistical data.

Geoestatistics

Let's suppose that $Y(\boldsymbol{s}_1), \ldots, Y(\boldsymbol{s}_n)$ are observations of a variable, Y is the measure of that variable and $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_n$ are the geographical location (e.g. latitude-longitude). Generally we assume that this is a realization of a stochastic process as:

$$\{Y(\boldsymbol{s}) : \boldsymbol{s} \in D \subset \mathbb{R}^2\}, \tag{2}$$

where D is a fixed subset in $\mathbb{R}^2$ Euclidean space.

For situations where $d > 1$, the process is referred to as *spatial process* or *random field*

Let's suppose that $Y(\boldsymbol{s}_1), \ldots, Y(\boldsymbol{s}_n)$ are observations of a variable, Y is the measure of that variable and $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_n$ are the geographical location (e.g. latitude-longitude). Generally we assume that this is a realization of a stochastic process as:

$$\{Y(\boldsymbol{s}) : \boldsymbol{s} \in D \subset \mathbb{R}^2\}, \tag{2}$$

where D is a fixed subset in $\mathbb{R}^2$ Euclidean space.

For situations where $d > 1$, the process is referred to as *spatial process* or *random field*

A Gaussian random field (GRF) is a sequence of random variables, where the observations come from a continuous space and have joint multivariate Normal distribution. This sequence of variables can be written as $\{Y(\boldsymbol{s}) : \boldsymbol{s} \in D \subset \mathbb{R}^2\}$, besides, this GRF can present the following characteristics:

- Stationarity
- Isotropy

For the above, the GRF has a mean (cte)

$$E[Y(\boldsymbol{s})] = \mu, \forall\ \boldsymbol{s} \in D \tag{3}$$

and the covariance depends only of the difference between sites $\boldsymbol{s}$:

$$Cov(Y(\boldsymbol{s}), Y(\boldsymbol{s}+\boldsymbol{h})) = C(\boldsymbol{h}), \forall\ \boldsymbol{s} \in D, \forall\ \boldsymbol{h} \in \mathbb{R}^2 \tag{4}$$

The covariance matrix Σ of a GRF specifies the dependence structure between the referenced points. There are different covariance function for the GRF, for example, if we consider $\boldsymbol{s}_i$ and $\boldsymbol{s}_j \in \mathbb{R}^2$ we have:

Exponential

$$\text{Cov}(Y(\boldsymbol{s}_i), Y(\boldsymbol{s}_j)) = \sigma^2 exp(-\kappa||\boldsymbol{s}_i - \boldsymbol{s}_j||) \tag{5}$$

where $||\boldsymbol{s}_i - \boldsymbol{s}_j||$ is the distance between $\boldsymbol{s}_i$ and $\boldsymbol{s}_j$, σ^2 is the variance of the random field and $\kappa > 0$ controls the correlation decay on function of the distance

Matérn

$$\text{Cov}(Y(\boldsymbol{s}_i), Y(\boldsymbol{s}_j)) = \frac{\sigma^2}{2^{\nu-1}\Gamma(\nu)}(\kappa||\boldsymbol{s}_i - \boldsymbol{s}_j||)^{\nu} K_{\nu}(\kappa||\boldsymbol{s}_i - \boldsymbol{s}_j||) \tag{6}$$

where σ^2 is the marginal variance of the random field, $K_{\nu}(\cdot)$ is the modified Bessel function and $\nu > 0$ is the smoothness parameter.

Classical spatial prediction

The classical approach for spatial prediction in Geostatistics is the *Kriging* (Matheron (1963)). For example; given observations of a random field $\boldsymbol{Y} = (Y(\boldsymbol{s}_1), \ldots Y(\boldsymbol{s}_n))^T$, how can we predict the variable Y in the site $\boldsymbol{s}^\star$? From a GRF perspective, consider a linear model where we have not covariates, we have only $Y(\boldsymbol{s}_i)$, so we can propose the following:

$$\boldsymbol{Y} = \mu \mathbf{1} + \varepsilon, \quad \text{where} \quad \varepsilon \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}) \tag{7}$$

For a spatial covariance (without nugget effect), we can write:

$$\boldsymbol{\Sigma} = \sigma^2 H(\phi), \quad \text{where} \quad H(\phi)_{ij} = \rho(\phi; d_{ij}) \tag{8}$$

$d_{ij} = |\boldsymbol{s}_i - \boldsymbol{s}_j|$ is the distance between $\boldsymbol{s}_i$ and $\boldsymbol{s}_j$ and ρ is a correlation function on $\mathbb{R}^d$. For a model with nugget effect we can write:

$$\boldsymbol{\Sigma} = \sigma^2 H(\phi) + \tau I \tag{9}$$

If we have covariates $\boldsymbol{x} = (x(\boldsymbol{s}_1) \ldots x(\boldsymbol{s}_n))^T$ the model has now a more general form:

$$\boldsymbol{Y} = \boldsymbol{X}\beta + \varepsilon, \quad \text{where } \varepsilon \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}) \tag{10}$$

,
then we can do:

- Make inference about the estimated parameters
- Predict at $\boldsymbol{s}^\star$ that we did not observed (Kriging)

If we have covariates $\boldsymbol{x} = (x(\boldsymbol{s}_1)\dots x(\boldsymbol{s}_n))^T$ the model has now a more general form:

$$\boldsymbol{Y} = \boldsymbol{X}\beta + \varepsilon, \text{ where } \varepsilon \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}) \tag{10}$$

,
then we can do:

- Make inference about the estimated parameters
- Predict at $\boldsymbol{s}^\star$ that we did not observed (Kriging)

But, what happen if our response variable (Y) has not a Normal distribution? Can we use the least square method to find the parameters in the linear spatial regression?

Beyond of the least squares method

Classical linear (Gaussian in the errors) model

Least square method

$$\boldsymbol{Y} = \boldsymbol{X}\beta + \varepsilon$$

with $\boldsymbol{\mu} = \boldsymbol{X}\beta$ and $\boldsymbol{Y} \sim \mathcal{N}(\boldsymbol{\mu}, \sigma^2)$. Thus $\hat{\beta} \sim \mathcal{N}(\beta, (\boldsymbol{X}^T\boldsymbol{X})^{-1}\sigma^2)$ and the least square estimation of β:

$$\mathcal{S} = \sum_{i=1}^{n}(Y_i - \mu)^2, \quad \text{with respect to} \quad \beta$$

This fitting follows from the log-likelihood for the Gaussian model (given the Normal assumption of ε) based on the Gauss Markov theorem.

Example: For a spatial Gaussian linear model

$$\begin{aligned} Y(\boldsymbol{s}) &= \mu(\boldsymbol{s}) + \epsilon(\boldsymbol{s}) \\ &= \mu(\boldsymbol{s}) + u(\boldsymbol{s}) + \varepsilon(\boldsymbol{s}), \end{aligned}$$

where:

- $E[y(\boldsymbol{s})] = \mu(\boldsymbol{s}) = x(\boldsymbol{s})^T\beta$
- $\epsilon(\boldsymbol{s})$ is the zero mean stationary process
- $u(\boldsymbol{s})$ is a spatially correlated process (A GRF)
- $\varepsilon(\boldsymbol{s})$ is the measurement error (commonly assumed $\mathcal{N}(0, \sigma_\varepsilon^2)$).

Likelihood method for the spatial Gaussian linear model

$Y(\cdot)$ is a GRF with mean $\boldsymbol{\mu} = \boldsymbol{X}^T \beta$ and covariance function:

$$C[Y(\boldsymbol{s}_1), Y(\boldsymbol{s}_2)] = C(\boldsymbol{s}_1, \boldsymbol{s}_2)$$

So, for observations $\boldsymbol{Y} = (Y(\boldsymbol{s}_1) \dots Y(\boldsymbol{s}_n))^T$, the mean vector is $\boldsymbol{X}\beta$ and a $n \times n$ covariance matrix $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ with entries $\boldsymbol{\Sigma}(\boldsymbol{\theta})_{i,j} = C(\boldsymbol{s}_i, \boldsymbol{s}_j)$. Thus, the distribution of the response variable is:

$$\boldsymbol{Y} \sim MVN(\boldsymbol{X}\beta, \boldsymbol{\Sigma}(\boldsymbol{\theta}))$$

The likelihood function for β and $\boldsymbol{\theta}$:

$$\mathcal{L}(\beta, \boldsymbol{\theta}) = (2\pi)^{-n/2} |\boldsymbol{\Sigma}(\boldsymbol{\theta})|^{-1/2} \exp\left\{ -\frac{1}{2}(\boldsymbol{Y} - \boldsymbol{X}\beta)^T \boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1} (\boldsymbol{Y} - \boldsymbol{X}\beta) \right\} \tag{11}$$

and the log-likelihood

$$\ell(\beta, \boldsymbol{\theta}) = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\log|\boldsymbol{\Sigma}(\boldsymbol{\theta})| - \frac{1}{2}(\boldsymbol{Y} - \boldsymbol{X}\beta)^T \boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1} (\boldsymbol{Y} - \boldsymbol{X}\beta) \tag{12}$$

Problem 1: The closed form (12) not always is obtained, and we typically turn to numerical optimization techniques

On the other hand, we are aware that:

- A Gaussian random field (GRF) is the main component of spatial modelling, but.....

- Another problem arises when we want to evaluate non Gaussian likelihood with a dense covariance matrix (Σ)

On the other hand, we are aware that:

- A Gaussian random field (GRF) is the main component of spatial modelling, but.....
- Another problem arises when we want to evaluate non Gaussian likelihood with a dense covariance matrix ($\boldsymbol{\Sigma}$)

If we use the classical definition of a GRF, then we need to consider:

$$\begin{pmatrix} u(\boldsymbol{s}_1) \\ u(\boldsymbol{s}_2) \\ \vdots \\ u(\boldsymbol{s}_n) \end{pmatrix} \sim N\left(\begin{pmatrix} \mu(\boldsymbol{s}_1) \\ \mu(\boldsymbol{s}_2) \\ \vdots \\ \mu(\boldsymbol{s}_n) \end{pmatrix}, \begin{pmatrix} c(\boldsymbol{s}_1, \boldsymbol{s}_1) & c(\boldsymbol{s}_1, \boldsymbol{s}_2) & \cdots & c(\boldsymbol{s}_1, \boldsymbol{s}_n) \\ c(\boldsymbol{s}_2, \boldsymbol{s}_1) & c(\boldsymbol{s}_2, \boldsymbol{s}_2) & \cdots & c(\boldsymbol{s}_2, \boldsymbol{s}_n) \\ \vdots & \vdots & \ddots & \vdots \\ c(\boldsymbol{s}_n, \boldsymbol{s}_1) & c(\boldsymbol{s}_n, \boldsymbol{s}_2) & \cdots & c(\boldsymbol{s}_n, \boldsymbol{s}_n) \end{pmatrix} \right), \tag{13}$$

where $\boldsymbol{s}_1, ..., \boldsymbol{s}_n$ are all of the distinct values of $\boldsymbol{s}_i$ in our spatial data.

Problem 2: Using the classical definition of a GRF, then:

- The storage scales quadratically in "n"
- The computation scales cubically in "n"

Problem 2: Using the classical definition of a GRF, then:

- The storage scales quadratically in "n"
- The computation scales cubically in "n"

Integrated Nested Laplace Approximation (INLA)

Commonly we have two paradigms for statistical modelling, for example:

Consider the following: $\boldsymbol{Y}$ is a set of observations with distribution of probability $\pi(\boldsymbol{Y} \mid \boldsymbol{\theta})$. For the above we can estimate $\boldsymbol{\theta}$ of two ways:

Frequentist approach

$\boldsymbol{\theta}$ denotes fixed and unknown parameters what can be estimated by maximum likelihood.

Bayesian approach

$\boldsymbol{\theta}$ denotes random variables with a prior $\pi(\boldsymbol{\theta})$ specification. We can estimate $\boldsymbol{\theta}$ based on the posterior:

$$\pi(\boldsymbol{\theta} \mid \boldsymbol{Y}) = \frac{\pi(\boldsymbol{Y} \mid \boldsymbol{\theta})\pi(\boldsymbol{\theta})}{\pi(\boldsymbol{Y})} \propto \pi(\boldsymbol{Y} \mid \boldsymbol{\theta})\pi(\boldsymbol{\theta}) \tag{14}$$

Specifically, in the Bayesian framework we can use:

- Hierarchical models to consider complex structures and explain the behavior of our data
- Propose a model to calculate the uncertainty associated with the parameters and latent variables (random effects)

How to do Bayesian inference in R?

R-INLA

The Integrated Nested Laplace Approximation (INLA) is a very used technique for spatial modelling available in R

This method was proposed by Rue et al. (2009). In summary, we can obtain the posterior distribution using numerical approximations. Advantage?

We don't need to do sampling

How to do Bayesian inference in R?

R-INLA

The Integrated Nested Laplace Approximation (INLA) is a very used technique for spatial modelling available in R

This method was proposed by Rue et al. (2009). In summary, we can obtain the posterior distribution using numerical approximations. Advantage?

We don't need to do sampling

How to do Bayesian inference in R?

R-INLA

The Integrated Nested Laplace Approximation (INLA) is a very used technique for spatial modelling available in R

This method was proposed by Rue et al. (2009). In summary, we can obtain the posterior distribution using numerical approximations. Advantage?

We don't need to do sampling

How to do Bayesian inference in R?

R-INLA

The Integrated Nested Laplace Approximation (INLA) is a very used technique for spatial modelling available in R

This method was proposed by Rue et al. (2009). In summary, we can obtain the posterior distribution using numerical approximations. Advantage?

We don't need to do sampling

How to do Bayesian inference in R?

R-INLA

The Integrated Nested Laplace Approximation (INLA) is a very used technique for spatial modelling available in R

This method was proposed by Rue et al. (2009). In summary, we can obtain the posterior distribution using numerical approximations. Advantage?

We don't need to do sampling

INLA works with a types of models called Latent Gaussian models. So, what is the idea behind of Latent Gaussian Models?

For example: Multiple linear regression model

$$\mu_i = \mathbb{E}(Y_i) = \beta_0 + \sum_{j=1}^{n_\beta} \beta_j x_{ji}, \quad i = 1, ..., n$$

where β_0 is the intercept and β are the parameters related to the covariates $\boldsymbol{x}$.

INLA works with a types of models called Latent Gaussian models. So, what is the idea behind of Latent Gaussian Models?

For example: Multiple linear regression model

$$\mu_i = \mathbb{E}(Y_i) = \beta_0 + \sum_{j=1}^{n_\beta} \beta_j x_{ji}, \quad i = 1, ..., n$$

where β_0 is the intercept and $\boldsymbol{\beta}$ are the parameters related to the covariates $\boldsymbol{x}$.

Idea behind of a Latent Gaussian Model

Generalized additive model (GAM)

$$\eta_i = g(\mu_i) = \beta_0 + \sum_{k=1}^{n_f} f_k(c_{ki}), \quad i = 1, ..., n$$

where $g(\cdot)$ es a link function, β_0 is the intercept, $f_k(\cdot)$ is the non-linear smopoth effects of the covariates $\boldsymbol{c}_k$.

Idea behind of a Latent Gaussian Model

A more complete general structure

$$\eta_i = g(\mu_i) = \beta_0 + \sum_{j=1}^{n_\beta} \beta_j x_{ji} + \sum_{k=1}^{n_f} f_k(c_{ki}), \quad i = 1, ..., n$$

where $g(\cdot)$ es a link function, β_0 is the intercept related to the covariates $\boldsymbol{x}$, $f_k(\cdot)$ is the non-linear smopoth effects of the covariates $\boldsymbol{c}_k$.

Latent Gaussian Models

So, we collect all the parameters of the linear predictor in a latent field

$$\boldsymbol{u} = \{\beta_0, \boldsymbol{\beta}, \{f_k(\cdot)\}, \boldsymbol{\eta}\}$$

and, in this way, we can assign a Gaussian prior (Essentially, a GMRF prior) to all the elements of $\boldsymbol{u}$.

A general way to express a model in INLA:

$$\eta_i = \beta_0 + \sum_{k=1}^{K} \beta_k x_{ki} + \sum_{l=1}^{L} f_l(z_{li})$$

where:

- β_0 is the intercept
- $(\beta_1 \dots \beta_K)$ are coefficients associated with the covariates $\boldsymbol{x} = ({}_1 \dots x_K)$
- $\boldsymbol{f} = (f(\cdot)_1 \dots f(\cdot)_L)$ is a set of functions defined and associated with some covariates $\boldsymbol{z} = ({}_1 \dots z_L)$

Finally:

$$\boldsymbol{\theta} = (\beta, \boldsymbol{f}) \sim \text{GMRF}(\boldsymbol{0}, \boldsymbol{Q})$$

Types of models that we can use with `INLA`:

- Generalized Linear Models (GLM)
- Generalized Linear Mixed Models (GLMM)
- Time series models
- Spatial models
- Spatio-temporal models

Again, let's review the assumption of a GRF..

$$\begin{pmatrix} u(\boldsymbol{s}_1) \\ u(\boldsymbol{s}_2) \\ \vdots \\ u(\boldsymbol{s}_n) \end{pmatrix} \sim N\left(\begin{pmatrix} \mu(\boldsymbol{s}_1) \\ \mu(\boldsymbol{s}_2) \\ \vdots \\ \mu(\boldsymbol{s}_n) \end{pmatrix}, \begin{pmatrix} c(\boldsymbol{s}_1, \boldsymbol{s}_1) & c(\boldsymbol{s}_1, \boldsymbol{s}_2) & \cdots & c(\boldsymbol{s}_1, \boldsymbol{s}_n) \\ c(\boldsymbol{s}_2, \boldsymbol{s}_1) & c(\boldsymbol{s}_2, \boldsymbol{s}_2) & \cdots & c(\boldsymbol{s}_2, \boldsymbol{s}_n) \\ \vdots & \vdots & \ddots & \vdots \\ c(\boldsymbol{s}_n, \boldsymbol{s}_1) & c(\boldsymbol{s}_n, \boldsymbol{s}_2) & \cdots & c(\boldsymbol{s}_n, \boldsymbol{s}_n) \end{pmatrix} \right),$$

Rue and Held (2005) approximates this problem assuming that the GRF is a Gaussian Markov random field (GMRF) $\rightarrow$ This means that the instead of to use the covariance matrix $\boldsymbol{\Sigma}$ we will use the precision matrix $\boldsymbol{Q} = \boldsymbol{\Sigma}^{-1}$

Universidad de Valparaíso CHILE

Again, let's review the assumption of a GRF..

$$\begin{pmatrix} u(\mathbf{s}_1) \\ u(\mathbf{s}_2) \\ \vdots \\ u(\mathbf{s}_n) \end{pmatrix} \sim N\left(\begin{pmatrix} \mu(\mathbf{s}_1) \\ \mu(\mathbf{s}_2) \\ \vdots \\ \mu(\mathbf{s}_n) \end{pmatrix}, \begin{pmatrix} c(\mathbf{s}_1,\mathbf{s}_1) & c(\mathbf{s}_1,\mathbf{s}_2) & \cdots & c(\mathbf{s}_1,\mathbf{s}_n) \\ c(\mathbf{s}_2,\mathbf{s}_1) & c(\mathbf{s}_2,\mathbf{s}_2) & \cdots & c(\mathbf{s}_2,\mathbf{s}_n) \\ \vdots & \vdots & \ddots & \vdots \\ c(\mathbf{s}_n,\mathbf{s}_1) & c(\mathbf{s}_n,\mathbf{s}_2) & \cdots & c(\mathbf{s}_n,\mathbf{s}_n) \end{pmatrix}\right),$$

Rue and Held (2005) approximates this problem assuming that the GRF is a Gaussian Markov random field (GMRF) $\rightarrow$ This means that the instead of to use the covariance matrix $\mathbf{\Sigma}$ we will use the precision matrix $\mathbf{Q} = \mathbf{\Sigma}^{-1}$

Again, let's review the assumption of a GRF..

$$\begin{pmatrix} u(\boldsymbol{s}_1) \\ u(\boldsymbol{s}_2) \\ \vdots \\ u(\boldsymbol{s}_n) \end{pmatrix} \sim N\left(\begin{pmatrix} \mu(\boldsymbol{s}_1) \\ \mu(\boldsymbol{s}_2) \\ \vdots \\ \mu(\boldsymbol{s}_n) \end{pmatrix}, \begin{pmatrix} c(\boldsymbol{s}_1, \boldsymbol{s}_1) & c(\boldsymbol{s}_1, \boldsymbol{s}_2) & \cdots & c(\boldsymbol{s}_1, \boldsymbol{s}_n) \\ c(\boldsymbol{s}_2, \boldsymbol{s}_1) & c(\boldsymbol{s}_2, \boldsymbol{s}_2) & \cdots & c(\boldsymbol{s}_2, \boldsymbol{s}_n) \\ \vdots & \vdots & \ddots & \vdots \\ c(\boldsymbol{s}_n, \boldsymbol{s}_1) & c(\boldsymbol{s}_n, \boldsymbol{s}_2) & \cdots & c(\boldsymbol{s}_n, \boldsymbol{s}_n) \end{pmatrix} \right),$$

Rue and Held (2005) approximates this problem assuming that the GRF is a Gaussian Markov random field (GMRF) $\rightarrow$ This means that the instead of to use the covariance matrix $\boldsymbol{\Sigma}$ we will use the precision matrix $\boldsymbol{Q} = \boldsymbol{\Sigma}^{-1}$

Considering the above, and taking advantage of the computational efficiency of using a GMRF, Lindgren et al. (2011) created an explicit link to approximate the GRF by a GMRF. This is:

$$C(\boldsymbol{s}_1, \boldsymbol{s}_2) = \frac{\sigma^2}{2^{\nu-1}\Gamma(\nu)}(\kappa||\boldsymbol{s}_2 - \boldsymbol{s}_1||)^{\nu} K_{\nu}(\kappa||\boldsymbol{s}_2 - \boldsymbol{s}_1||), \tag{15}$$

where $||\boldsymbol{s}_2 - \boldsymbol{s}_1||$ is the Euclidean distance between two geographical points $\boldsymbol{s}_1$ and $\boldsymbol{s}_2 \in \mathcal{R}^D$, K_ν is the modified Bessel function with $\nu > 0$, $\kappa > 0$ what controls the correlation through $\rho = \sqrt{8\nu}/\kappa$, and σ^2 is the marginal variance.

The authors noted that the GRF ($u(\boldsymbol{s})$) and Matérn function (15) has solution to the linear fractional SPDE

$$(\kappa^2 - \Delta)^{\alpha/2}(\tau u(\boldsymbol{s})) = W(\boldsymbol{s}), \quad \boldsymbol{s} \in \mathcal{R}^D, \quad \text{with} \quad \alpha = \nu + d/2, \quad \kappa > 0, \nu > 0, \tag{16}$$

where W is a spatial Gaussian white noise (Whittle (1954), Whittle (1963)), Δ is the Laplacian operator and τ controls the marginal variance as:

$$\tau^2 = \frac{\Gamma(\nu)}{\Gamma(\nu + d/2)(4\pi)^{d/2}\kappa^{2\nu}\sigma^2} \tag{17}$$

So, to find $u(\boldsymbol{s})$ with Matérn function (15) then is necessary to solve (16).

With additional other mathematical calculus, Lindgren et al. (2011) used the elements finite method to represent $u(\boldsymbol{s})$ in a non structured triangulation as:

$$u_h(\boldsymbol{s}) = \sum_{k=1}^{n} w_k \psi_k(\boldsymbol{s}), \quad (18)$$

where $\{\psi_k\}_{k=1}^{n}$ are piecewise linear basis functions. Finally, they showed that the Gaussian coefficients $\{w_k\}_{k=1}^{n}$ are GMRF when $\alpha = 1$ and can be approximated with a GMRF when $\alpha = 2$ (Liu et al. (2016)).

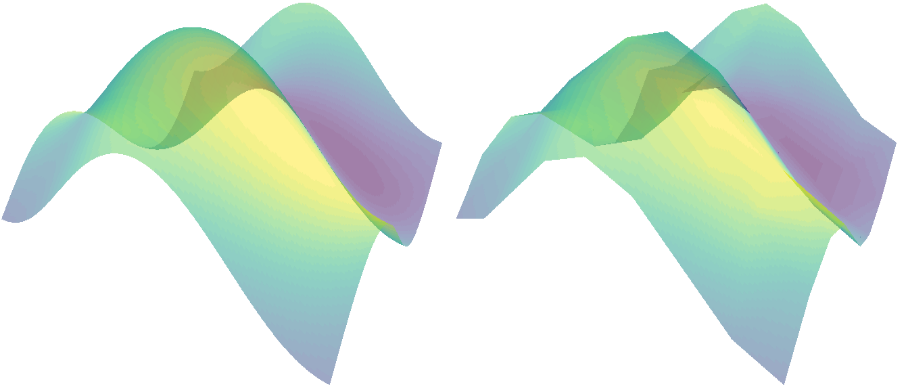


Figure 4: A GRF (left) approximated by a GMRF (right), (Krainski et al. (2018))

Lindgren et al. (2021) did a list with recent applications of the SPDE method in different areas of the research, for example:

- Astronomy (Levis et al. (2021))
- Health (Moraga et al. (2021), INLA et al. (2021))
- Engineering (Zhang et al. (2021))
- Theory (Ghattas and Willcox (2021))
- Environmetrics (Hough et al. (2021))
- Imaging (Aquino et al. (2021))
- Fisheries (Cavieres et al. (2021))
- .....

More of this references in Lindgren et al. (2021)

How we can use the approximate GRF ($\sim$ GMRF) in a Bayesian spatial (spatio-temporal) model?

A very quick explanation of how works INLA in the spatial context:

- The GRF is parameterized by the precision matrix $\boldsymbol{Q} = \boldsymbol{\Sigma}^{-1}$.
- We don't built a discrete model for the GRF on a grid, we construct an approximation of the GRF in a spatial continuous space defined on the entire study area.
- INLA done the inference for univariate posterior densities for the parameters of $u(\boldsymbol{s})$, and the joint posterior of the hyperparameters of the model.

So, how can we express a spatial model in INLA? → Hierarchical model!

$$\begin{aligned}
\boldsymbol{\theta} &\sim \boldsymbol{\theta} && \text{Hyperparameters} && (19)\\
\boldsymbol{u} \mid \boldsymbol{\theta} &\sim \mathcal{N}(\mathbf{0}, \boldsymbol{Q}(\boldsymbol{\theta})^{-1}) && \text{Latent Gaussian field} && (20)\\
y_i \mid \boldsymbol{u}, \boldsymbol{\theta} &\sim \prod_i \pi(y_i \mid \eta_i, \boldsymbol{\theta}) && \text{Observations} && (21)
\end{aligned}$$

where $\boldsymbol{Q}(\boldsymbol{\theta})$ is the precision matrix, $\boldsymbol{u}$ is the latent Gaussian field and $\eta_i = log(\mu_i) = \text{intercept} + f(\boldsymbol{X}_i) + \boldsymbol{u}_i$, where the matrix $\boldsymbol{X}$ is a set of covariates and $\boldsymbol{u} \sim \text{GMRF}(0, \boldsymbol{Q}^{-1})$

So, how can we express a spatial model in `INLA`? $\rightarrow$ Hierarchical model!

$$\boldsymbol{\theta} \sim \boldsymbol{\theta} \quad \text{Hyperparameters} \quad (19)$$

$$\boldsymbol{u} \mid \boldsymbol{\theta} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{Q}(\boldsymbol{\theta})^{-1}) \quad \text{Latent Gaussian field} \quad (20)$$

$$y_i \mid \boldsymbol{u}, \boldsymbol{\theta} \sim \prod_i \pi(y_i \mid \eta_i, \boldsymbol{\theta}) \quad \text{Observations} \quad (21)$$

where $\boldsymbol{Q}(\boldsymbol{\theta})$ is the precision matrix, $\boldsymbol{u}$ is the latent Gaussian field and $\eta_i = log(\mu_i) = \text{intercept} + f(\boldsymbol{X}_i) + \boldsymbol{u}_i$, where the matrix $\boldsymbol{X}$ is a set of covariates and $\boldsymbol{u} \sim \text{GMRF}(0, \boldsymbol{Q}^{-1})$

So, how can we express a spatial model in INLA? → Hierarchical model!

$$
\begin{aligned}
\boldsymbol{\theta} &\sim \boldsymbol{\theta} && \text{Hyperparameters} && (19)\\
\boldsymbol{u} \mid \boldsymbol{\theta} &\sim \mathcal{N}(\mathbf{0}, \boldsymbol{Q}(\boldsymbol{\theta})^{-1}) && \text{Latent Gaussian field} && (20)\\
y_i \mid \boldsymbol{u}, \boldsymbol{\theta} &\sim \prod_i \pi(y_i \mid \eta_i, \boldsymbol{\theta}) && \text{Observations} && (21)
\end{aligned}
$$

where $\boldsymbol{Q}(\boldsymbol{\theta})$ is the precision matrix, $\boldsymbol{u}$ is the latent Gaussian field and $\eta_i = log(\mu_i) = \text{intercept} + f(\boldsymbol{X}_i) + \boldsymbol{u}_i$, where the matrix $\boldsymbol{X}$ is a set of covariates and $\boldsymbol{u} \sim \text{GMRF}(0, \boldsymbol{Q}^{-1})$

Universidad de Valparaíso CHILE

Examples

Conclusions

- `INLA` is an efficient tool to estimate different statistical models.
- The estimation is faster than MCMC method for Bayesian Inference.
- Is an excellent alternative to fit geostatistical spatial/spatio-temporal models based on the SPDE method.

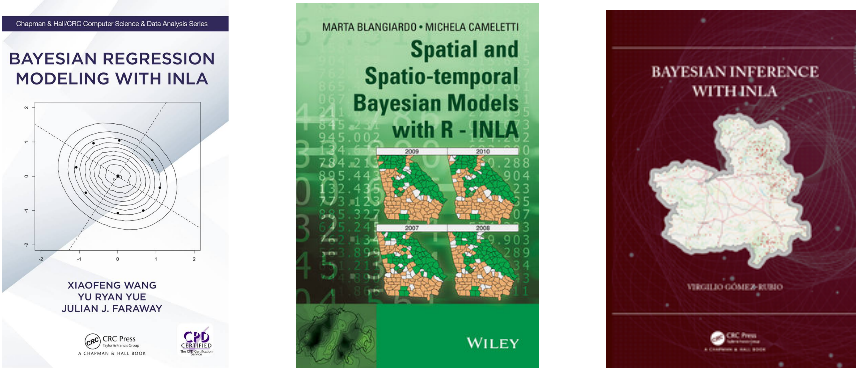


Figure 5: Some books to learn about INLA

Thank You

References I

Aquino, B., Castruccio, S., Gupta, V., and Howard, S. (2021). Spatial modeling of mid-infrared spectral data with thermal compensation using integrated nested laplace approximation. *Applied Optics*, 60(27):8609–8615.

Cavieres, J., Monnahan, C. C., and Vehtari, A. (2021). Accounting for spatial dependence improves relative abundance estimates in a benthic marine species structured as a metapopulation. *Fisheries Research*, 240:105960.

Cressie, N. A. (1993). Spatial prediction and kriging. *Statistics for Spatial Data (Cressie NAC, ed). New York: John Wiley & Sons*, pages 105–209.

Ghattas, O. and Willcox, K. (2021). Learning physics-based models from data: perspectives from inverse problems and model reduction. *Acta Numerica*, 30:445–554.

References II

Hough, I., Sarafian, R., Shtein, A., Zhou, B., Lepeule, J., and Kloog, I. (2021). Gaussian markov random fields improve ensemble predictions of daily 1 km pm2. 5 and pm10 across france. *Atmospheric Environment*, 264:118693.

INLA, U. D. L., LESTIMATION, S. P., and ALGERIE, P. E. (2021). Using inla/spde approach for estimating a spatial model for lung cancer mortality in algeria 2016. *Revue dEconomie et de Statistique Appliquée*, 18(1).

Krainski, E., Gómez-Rubio, V., Bakka, H., Lenzi, A., Castro-Camilo, D., Simpson, D., Lindgren, F., and Rue, H. (2018). *Advanced spatial modeling with stochastic partial differential equations using R and INLA*. Chapman and Hall/CRC.

References III

Levis, A., Lee, D., Tropp, J. A., Gammie, C. F., and Bouman, K. L. (2021). Inference of black hole fluid-dynamics from sparse interferometric measurements. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 2340–2349.

Lindgren, F., Bolin, D., and Rue, H. (2021). The spde approach for gaussian and non-gaussian fields: 10 years and still running. *arXiv preprint arXiv:2111.01084*.

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):423–498.

Liu, X., Guillas, S., and Lai, M.-J. (2016). Efficient spatial modeling using the spde approach with bivariate splines. *Journal of Computational and Graphical Statistics*, 25(4):1176–1194.

References IV

Matheron, G. (1963). Principles of geostatistics. *Economic geology*, 58(8):1246–1266.

Moraga, P. (2019). *Geospatial Health Data: Modeling and Visualization with R-INLA and Shiny*. CRC Press.

Moraga, P., Dean, C., Inoue, J., Morawiecki, P., Noureen, S. R., and Wang, F. (2021). Bayesian spatial modelling of geostatistical data using inla and spde methods: A case study predicting malaria risk in mozambique. *Spatial and Spatio-temporal Epidemiology*, 39:100440.

Pebesma, E. (2018). sf: Simple features for r. *R package version 0.6-0*.

Rue, H. and Held, L. (2005). *Gaussian Markov random fields: theory and applications*. CRC press.

References V

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.

Whittle, P. (1954). On stationary processes in the plane. *Biometrika*, pages 434–449.

Whittle, P. (1963). Stochastic-processes in several dimensions. *Bulletin of the International Statistical Institute*, 40(2):974–994.

Zhang, H., Guilleminot, J., and Gomez, L. J. (2021). Stochastic modeling of geometrical uncertainties on complex domains, with application to additive manufacturing and brain interface geometries. *Computer methods in applied mechanics and engineering*, 385:114014.