Spatial and spatio-temporal statistics for ecology

B. Alglave

Décembre 2022

Section 1

Introduction

Introduction

All ecological processes have a spatial and a spatio-temporal dimension:

- species distribution vary along the year following the season, the life stage, etc.

- ecosystem productivity is heterogeneous in space and time

- climatic variables evolve in response to climate change

➠ There is a need to develop mathematical tools to study/understand/model these processes

Introduction

All ecological processes have a spatial and a spatio-temporal dimension:

- species distribution vary along the year following the season, the life stage, etc.

- ecosystem productivity is heterogeneous in space and time

- climatic variables evolve in response to climate change

⟼ **There is a need to develop mathematical tools to study/understand/model these processes**

Introduction

All ecological processes have a spatial and a spatio-temporal dimension:

- species distribution vary along the year following the season, the life stage, etc.

- ecosystem productivity is heterogeneous in space and time

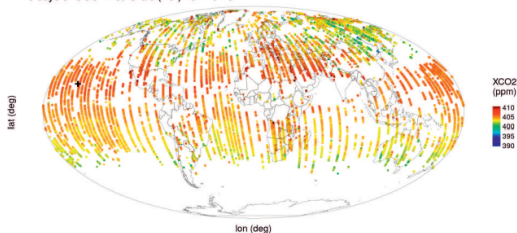- climatic variables evolve in response to climate change

➠ **There is a need to develop mathematical tools to study/understand/model these processes**

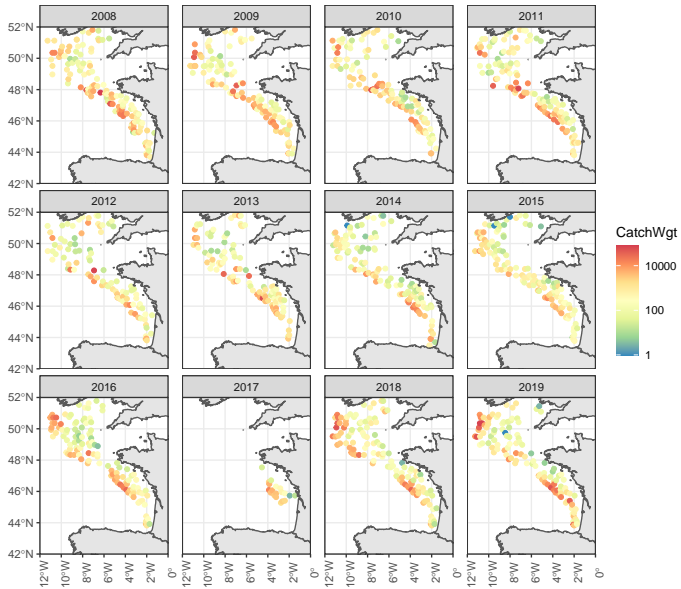**Examples:** what are the ecological processes we would like to infer from these data?

16 days of OCO-2 retrievals (V8r) 2017-01-01
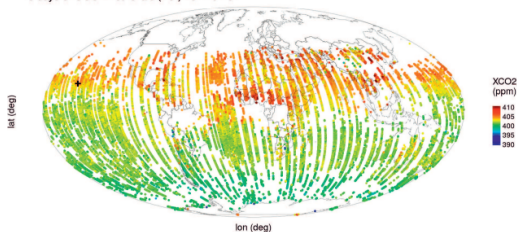
16 days of OCO-2 retrievals (V8r) 2017-07-01

Argentina sphyraena (EVHOE)

Data and ecological processes

What are the characteristics of these data?

- They arise from the ecological process of interest (= they are **conditionnal** on the ecological process)
- **Noisy** (= they are not perfect observations of the ecological process)
- **Sparse** (= they do not cover the full area and some hypothesis need to be set to predict the ecological process between the sampled locations)

What are the characteristics of the ecological process we want to infer?

- **Hidden** or **latent**
- **Structured** (relations that structure the process?)

How to relate the data to the ecological process?

➠ **Hierarchical modeling**

Data and ecological processes

What are the characteristics of these data?
- They arise from the ecological process of interest (= they are **conditionnal** on the ecological process)
- **Noisy** (= they are not perfect observations of the ecological process)
- **Sparse** (= they do not cover the full area and some hypothesis need to be set to predict the ecological process between the sampled locations)

What are the characteristics of the ecological process we want to infer?
- **Hidden** or **latent**
- **Structured** (relations that structure the process?)

How to relate the data to the ecological process?

➠ **Hierarchical modeling**

Data and ecological processes

What are the characteristics of these data?
- They arise from the ecological process of interest (= they are **conditionnal** on the ecological process)
- **Noisy** (= they are not perfect observations of the ecological process)
- **Sparse** (= they do not cover the full area and some hypothesis need to be set to predict the ecological process between the sampled locations)

What are the characteristics of the ecological process we want to infer?
- **Hidden** or **latent**
- **Structured** (relations that structure the process?)

How to relate the data to the ecological process?

➠ **Hierarchical modeling**

Data and ecological processes

What are the characteristics of these data?

- They arise from the ecological process of interest (= they are **conditionnal** on the ecological process)
- **Noisy** (= they are not perfect observations of the ecological process)
- **Sparse** (= they do not cover the full area and some hypothesis need to be set to predict the ecological process between the sampled locations)
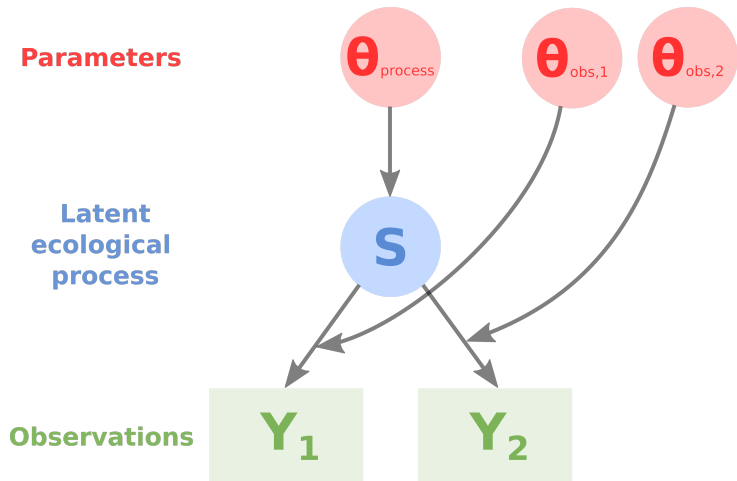
What are the characteristics of the ecological process we want to infer?

- **Hidden** or **latent**
- **Structured** (relations that structure the process?)

How to relate the data to the ecological process?

➠ **Hierarchical modeling**

Hierarchical models

Parameters

$\theta_{process}$ $\theta_{obs,1}$ $\theta_{obs,2}$

Latent ecological process

S

Observations

Y_1 Y_2

Observation process

$$\mathbf{Y}|\mathbf{S}, \boldsymbol{\theta}_{obs} \sim \mathcal{L}(\mathbf{S}, \boldsymbol{\theta}_{obs})$$

Y: observations
S: latent field
$\mathcal{L}$: probability distribution of **Y**
$\boldsymbol{\theta}_{obs}$: observation parameters

Hidden process

$$\mathbf{S}|\boldsymbol{\theta}_{process} \sim \mathcal{F}(\boldsymbol{\theta}_{process})$$

$\mathcal{F}$: probability distribution of **S**
$\theta_{process}$: process parameters

➠ **How to specify such model for space-time applications?**
➠ **What are the specificities of spatial and spatio-temporal models?**

Observation process

$$\mathbf{Y}|\mathbf{S}, \boldsymbol{\theta}_{obs} \sim \mathcal{L}(\mathbf{S}, \boldsymbol{\theta}_{obs})$$

$\mathbf{Y}$: observations
$\mathbf{S}$: latent field
$\mathcal{L}$: probability distribution of $\mathbf{Y}$
$\boldsymbol{\theta}_{obs}$: observation parameters

Hidden process

$$\mathbf{S}|\boldsymbol{\theta}_{process} \sim \mathcal{F}(\boldsymbol{\theta}_{process})$$

$\mathcal{F}$: probability distribution of $\mathbf{S}$
$\boldsymbol{\theta}_{process}$: process parameters

➠ **How to specify such model for space-time applications?**
➠ **What are the specificities of spatial and spatio-temporal models?**

Section 2

Basis of spatio-temporal statistical modeling

Basis of spatio-temporal statistical modeling

Linear models are basis of statistical models.

For instance, in the case of EVHOE data for Argentina sphyraena, let's assume that species spatio-temporal distribution **S** vary in space (x), time (t) and depends on a quadratic effect of depth.

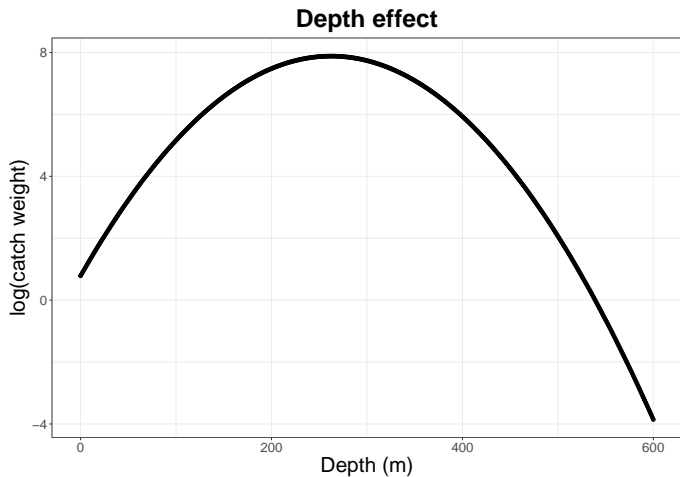$$\log(Y_i) \overset{i.i.d}{\sim} \mathcal{N}(S(x_i, t_i), \sigma^2)$$

$$S(x, t) = \mu + \beta_1 \cdot \text{Depth}(x) + \beta_2 \cdot (\text{Depth}(x))^2$$

or alternatively,

$$\log(Y_i) = \mu + \beta_1 \cdot \text{Depth}(x_i) + \beta_2 \cdot (\text{Depth}(x_i)) + \epsilon_i$$
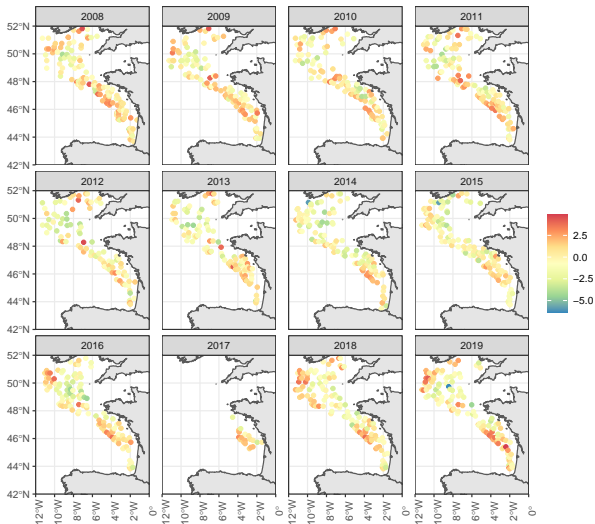
$$\epsilon_i \overset{i.i.d}{\sim} \mathcal{N}(0, \sigma^2)$$

Depth effect

However, the i.i.d hypothesis is rarely satisfied in a spatio-temporal context as there can be spatio-temporal correlation structures in the residuals.

Residuals of the linear model

Accounting for spatial and spatio-temporal dependencies

➠ Introduce a Gaussian field ($\mathcal{GF}$) δ in the hierarchical framework that captures spatial / spatio-temporal correlation

$$\log(Y_i) \overset{i.i.d}{\sim} \mathcal{N}(S(x_i, t_i), \sigma^2)$$

$$S(x, t) = \mu + \beta_1 \cdot \text{Depth}(x) + \beta_2 \cdot \text{Depth}(x)^2 + \delta(x, t)$$
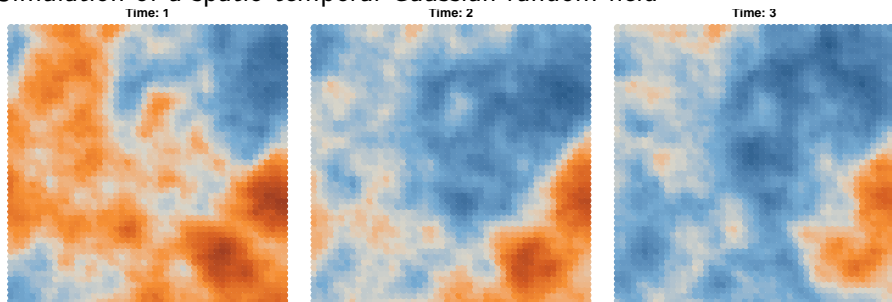
$$\delta \sim \mathcal{GF}(0, \mathcal{C}(x, y; t, r))$$

with $\mathcal{C}(x, y; t, r)$ the spatio-temporal covariance function that controls the structure of δ

Defining a Gaussian Field

What is a Gaussian field?

Simulation of a spatio-temporal Gaussian random field

Probability density function of a multivariate Normal distribution:

$$f_{MN}(\delta_1, ..., \delta_n) = (2\pi)^{-n/2} |^\bullet|^{-1/2} \exp\left(-\frac{1}{2}\boldsymbol{\delta}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\delta}\right)$$

What is a covariance function?

Variance

For a random variable X:

$$\text{Var}(X) = \text{E}\left[(X - \text{E}[X])^2\right]$$
$$= \text{E}\left[X^2\right] - \text{E}[X]^2$$

➠ Measure of dispersion. How far a set of numbers is spread out from their average value?
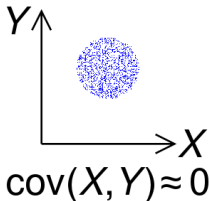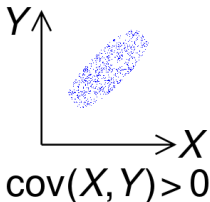
Covariance

We define 2 random variables X and Y.

$$\text{Cov}(X, Y) \equiv \text{E}[(X - \text{E}[X])(Y - \text{E}[Y])]$$

➠ Measure of the joint variability of two random variables.

The **correlation coefficient** is a normalized measure of the covariance.

$$\rho_{X,Y} = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y}$$

Y

X

$\text{cov}(X,Y) < 0$

Y

X

$\text{cov}(X,Y) \approx 0$

Y

X

$\text{cov}(X,Y) > 0$

Variance

For a random variable X:

$$\text{Var}(X) = \text{E}\left[(X - \text{E}[X])^2\right]$$
$$= \text{E}\left[X^2\right] - \text{E}[X]^2$$

➠ Measure of dispersion. How far a set of numbers is spread out from their average value?
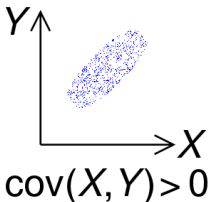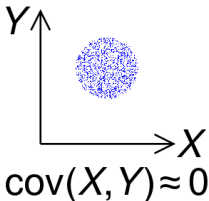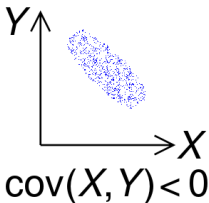
Covariance

We define 2 random variables X and Y.

$$\text{Cov}(X, Y) \equiv \text{E}[(X - \text{E}[X])(Y - \text{E}[Y])]$$

➠ Measure of the joint variability of two random variables.

The **correlation coefficient** is a normalized measure of the covariance.

$$\rho_{X,Y} = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y}$$

$\text{cov}(X, Y) < 0$

$\text{cov}(X, Y) \approx 0$

$\text{cov}(X, Y) > 0$

Variance

For a random variable X:

$$\text{Var}(X) = \text{E}\left[(X - \text{E}[X])^2\right]$$
$$= \text{E}\left[X^2\right] - \text{E}[X]^2$$

➠ Measure of dispersion. How far a set of numbers is spread out from their average value?
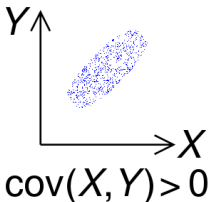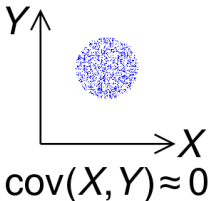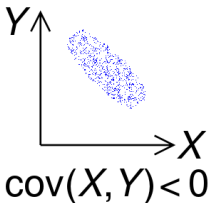
Covariance

We define 2 random variables X and Y.

$$\text{Cov}(X, Y) \equiv \text{E}[(X - \text{E}[X])(Y - \text{E}[Y])]$$

➠ Measure of the joint variability of two random variables.

The **correlation coefficient** is a normalized measure of the covariance.

$$\rho_{X,Y} = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y}$$

cov$(X, Y) < 0$

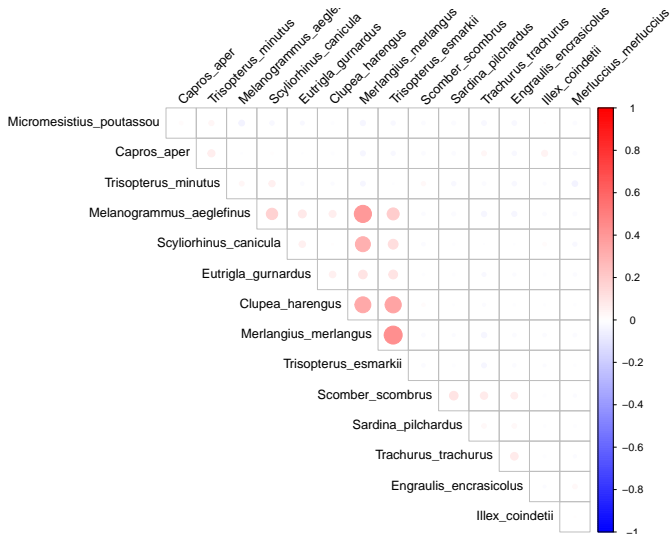cov$(X, Y) \approx 0$

cov$(X, Y) > 0$

Variance-covariance matrix

For a vector of p random variables $\mathbf{X} = (X_1, ..., X_p)$:

$$\mathrm{Var}(\mathbf{X}) = \begin{pmatrix} \mathrm{Var}(X_1) & \mathrm{Cov}(X_1, X_2) & \cdots & \mathrm{Cov}(X_1, X_p) \\ \mathrm{Cov}(X_2, X_1) & \ddots & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{Cov}(X_p, X_1) & \cdots & \cdots & \mathrm{Var}(X_p) \end{pmatrix}$$

The correlation matrix is the normalized version of the covariance matrix.

$\mathrm{Var}(\mathbf{X})^{-1}$ is the precision matrix.

Correlation matrix for the catch weights
of the 15 most abundant species of EVHOE data

Spatial variance-covariance

Let's consider the random effect $\delta(x, t)$ at a given time step t that we rewrite $\delta(x)$.

The spatial variance-covariance matrix of $\delta(x)$ can be written:

$$
\mathsf{Var}(\boldsymbol{\delta}(.)) = \begin{pmatrix}
\mathsf{Var}\left(\delta(x_1)\right) & \mathsf{Cov}\left(\delta(x_1), \delta(x_2)\right) & \cdots & \mathsf{Cov}\left(\delta(x_1), \delta(x_n)\right) \\
\mathsf{Cov}\left(\delta(x_2), \delta(x_1)\right) & \ddots & \cdots & \vdots \\
\vdots & \vdots & \ddots & \vdots \\
\mathsf{Cov}\left(\delta(x_n), \delta(x_1)\right) & \cdots & \cdots & \mathsf{Var}\left(\delta(x_n)\right)
\end{pmatrix}
$$

➡ **No spatial representation of the covariance matrix per se**

The spatial dependencies between the values of $\boldsymbol{\delta}(.)$ is modeled through a variance-covariance function $\mathcal{C}(x, y)$ (spatial only here).

Some base properties are required to define a spatial covariance function.

Necessary properties for spatial covariance matrix

Non-negative and positive-definiteness

A function $\{\mathcal{C}(x, y) : x, y \in D\}$ defined on D is said to be **non-negative-definite**, if for any complex numbers $\{a_i : i = 1, ..., m\}$, any $\{x_i : i = 1, ..., m\}$ in D, and any integer m, we have

$$\sum_{i=1}^{m} \sum_{j=1}^{m} a_i \bar{a}_j \mathcal{C}(x_i, x_j) \geq 0$$

where $\bar{a}$ denotes the complex conjugate of a.

To be *valid*, a covariance function must be non-negative definite.

A function is **positive-definite** when the inequality below is strictly positive whenever $(a_1, ..., a_m)'$ is a nonzero vector.

Additionnal (convenient) properties for spatial covariance matrix

Stationarity

There are 2 main kinds of stationarity: strong and second-order (or weak) stationarity.

- $\delta(x)$ is strongly stationary when the two probability measures defining $\delta(x)$ and $\delta(x + h)$ are equivalent for all $h \in \mathbb{R}^d$.

- $\delta(x)$ is second-order (or weakly) stationary when it has a constant expectation and a stationary covariance function.

Isotropy

Spatial isotropy corresponds to $Cov(\delta(x), \delta(y)) \equiv \mathcal{C}(\|x - y\|)$

Intrinsic

$\delta(x)$ is intrinsic if for all $(x) \in \mathbb{R}^d$, $h \in \mathbb{R}$, $\delta(x + h) + \delta(x)$ is second-ordered stationary (constant expectation and stationary covariance).

Some examples of isotropic covariance functions

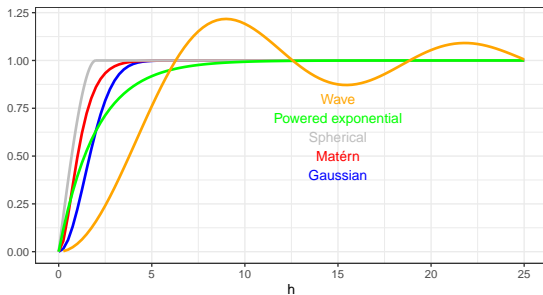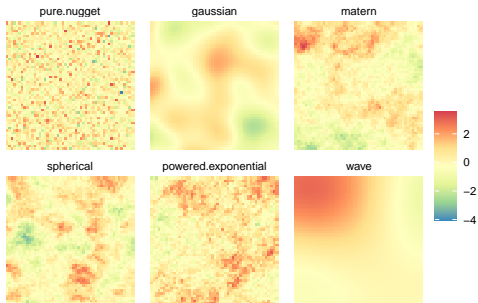| Model	Covariance function, $C(d)$
Linear	$C(d)$ does not exist
Spherical	$C(d) = \begin{cases} 0 & \text{if } d \geq 1/\phi \\ \sigma^2 \left[1 - \frac{3}{2}\phi d + \frac{1}{2}(\phi d)^3\right] & \text{if } 0 < d \leq 1/\phi \\ \tau^2 + \sigma^2 & d = 0 \end{cases}$
Exponential	$C(d) = \begin{cases} \sigma^2 \exp(-\phi d) & \text{if } d > 0 \\ \tau^2 + \sigma^2 & d = 0 \end{cases}$
Powered exponential	$C(d) = \begin{cases} \sigma^2 \exp(-|\phi d|^p) & \text{if } d > 0 \\ \tau^2 + \sigma^2 & d = 0 \end{cases}$
Gaussian	$C(d) = \begin{cases} \sigma^2 \exp(-\phi^2 d^2) & \text{if } d > 0 \\ \tau^2 + \sigma^2 & d = 0 \end{cases}$
Rational quadratic	$C(d) = \begin{cases} \sigma^2 \left(1 - \frac{d^2}{(\phi + d^2)}\right) & \text{if } d > 0 \\ \tau^2 + \sigma^2 & d = 0 \end{cases}$
Wave	$C(d) = \begin{cases} \sigma^2 \frac{\sin(\phi d)}{\phi d} & \text{if } d > 0 \\ \tau^2 + \sigma^2 & d = 0 \end{cases}$
Power law	$C(d)$ does not exist
Matérn	$C(d) = \begin{cases} \frac{\sigma^2}{2^{\nu-1}\Gamma(\nu)} (\phi d)^\nu K_\nu(\phi d) & \text{if } d > 0 \\ \tau^2 + \sigma^2 & d = 0 \end{cases}$
Matérn at $\nu = 3/2$	$C(d) = \begin{cases} \sigma^2 (1 + \phi d) \exp(-\phi d) & \text{if } d > 0 \\ \tau^2 + \sigma^2 & d = 0 \end{cases}$

Another (more general) way to define spatial structure is to define the **covariogram** $Var(Y(x) - Y(y)) = 2\gamma(x, y)$ (with $\gamma()$ the **semivariogram**) and its stationary version $2\gamma(h) = 2(\prime - \langle); h \in \mathbb{R}$.
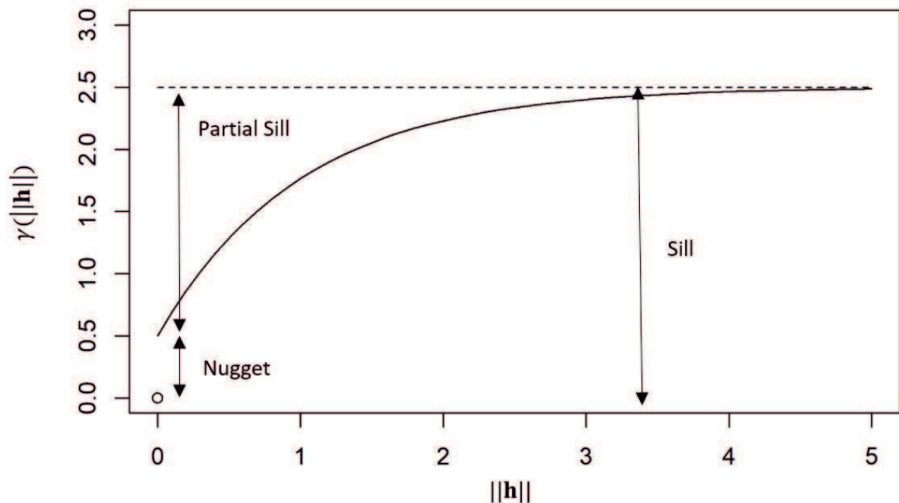
Some examples of isotropic variograms

Model	Variogram, $\gamma(d)$
Linear	$\gamma(d) = \begin{cases} \tau^2 + \sigma^2 d & \text{if } d > 0 \\ 0 & d = 0 \end{cases}$
Spherical	$\gamma(d) = \begin{cases} \tau^2 + \sigma^2 & \text{if } d \geq 1/\phi \\ \tau^2 + \sigma^2 \left[\frac{3}{2}\phi d - \frac{1}{2}(\phi d)^3\right] & \text{if } 0 < d \leq 1/\phi \\ 0 & d = 0 \end{cases}$
Exponential	$\gamma(d) = \begin{cases} \tau^2 + \sigma^2(1 - \exp(-\phi d)) & \text{if } d > 0 \\ 0 & d = 0 \end{cases}$
Powered exponential	$\gamma(d) = \begin{cases} \tau^2 + \sigma^2(1 - \exp(-|\phi d|^p)) & \text{if } d > 0 \\ 0 & d = 0 \end{cases}$
Gaussian	$\gamma(d) = \begin{cases} \tau^2 + \sigma^2(1 - \exp(-\phi^2 d^2)) & \text{if } d > 0 \\ 0 & d = 0 \end{cases}$
Rational quadratic	$\gamma(d) = \begin{cases} \tau^2 + \frac{\sigma^2 d^2}{(\phi + d^2)} & \text{if } d > 0 \\ 0 & d = 0 \end{cases}$
Wave	$\gamma(d) = \begin{cases} \tau^2 + \sigma^2(1 - \frac{\sin(\phi d)}{\phi d}) & \text{if } d > 0 \\ 0 & d = 0 \end{cases}$
Power law	$\gamma(d) = \begin{cases} \tau^2 + \sigma^2 d^\lambda & \text{if } d > 0 \\ 0 & d = 0 \end{cases}$
Matérn	$\gamma(d) = \begin{cases} \tau^2 + \sigma^2 \left[1 - \frac{(\phi d)^\nu}{2^{\nu-1}\Gamma(\nu)} K_\nu(\phi d)\right] & \text{if } d > 0 \\ 0 & d = 0 \end{cases}$
Matérn at $\nu = 3/2$	$\gamma(d) = \begin{cases} \tau^2 + \sigma^2 \left[1 - (1 + \phi d)\exp(-\phi d)\right] & \text{if } d > 0 \\ 0 & d = 0 \end{cases}$

Moving to spatio-temporal

The ideas are similar, but we add temporal correlations in the expression of the random effect δ.

For instance, let's introduce the model:

$$\delta(x, t) = \varphi \cdot \delta(x, t-1) + \omega(x, t) \text{ for } t = 2, ..., T$$

- $\varphi \in \,] -1 ; 1 [$ is the autoregressive temporal term
- $\omega(x, t)$ is a purely spatial GRF
- $\omega(x, 1)$ derives from the stationary distribution $\mathcal{N}(0, \sigma^2/(1 - \varphi^2))$

$$\mathcal{C}\left(\omega\left(x, t\right), \omega\left(y, r\right)\right) = \begin{cases} 0 & \text{if } t \neq r \\ \sigma_\omega^2 \cdot Cor(h) & \text{if } t = r \end{cases} \quad \text{for } x \neq y$$

where $h = ||x - y|| \in \mathbb{R}$ and $Cor(h)$ is the Matérn correlation function.

The variance-covariance matrix can be written as:

$$\mathcal{C}\left(\delta\left(x, t\right), \delta\left(y, r\right)\right) = \frac{\varphi^{|t-r|}}{1 - \varphi^2} \sigma_\omega^2 \, Cor(h)$$

with $|t - r|$ the time lag between time step t and r.

Moving to spatio-temporal

The ideas are similar, but we add temporal correlations in the expression of the random effect $\boldsymbol{\delta}$.

For instance, let's introduce the model:

$$\delta(x,t) = \varphi \cdot \delta(x, t-1) + \omega(x,t) \text{ for } t = 2, ..., T$$

- $\varphi \in]-1; 1[$ is the autoregressive temporal term
- $\omega(x,t)$ is a purely spatial GRF
- $\omega(x,1)$ derives from the stationary distribution $\mathcal{N}(0, \sigma^2/(1-\varphi^2))$

$$\mathcal{C}\left(\omega\left(x,t\right), \omega\left(y,r\right)\right) = \begin{cases} 0 & \text{if } t \neq r \\ \sigma_\omega^2 \cdot Cor(h) & \text{if } t = r \end{cases} \quad \text{for } x \neq y$$

where $h = ||x - y|| \in \mathbb{R}$ and $Cor(h)$ is the Matérn correlation function.

The variance-covariance matrix can be written as:

$$\mathcal{C}\left(\delta\left(x,t\right), \delta\left(y,r\right)\right) = \frac{\varphi^{|t-r|}}{1-\varphi^2}\sigma_\omega^2 Cor(h)$$

with $|t - r|$ the time lag between time step t and r.

Moving to spatio-temporal

The ideas are similar, but we add temporal correlations in the expression of the random effect δ.

For instance, let's introduce the model:

$$\delta(x, t) = \varphi \cdot \delta(x, t-1) + \omega(x, t) \text{ for } t = 2, ..., T$$

- $\varphi \in]-1; 1[$ is the autoregressive temporal term
- $\omega(x, t)$ is a purely spatial GRF
- $\omega(x, 1)$ derives from the stationary distribution $\mathcal{N}(0, \sigma^2/(1 - \varphi^2))$

$$\mathcal{C}(\omega(x, t), \omega(y, r)) = \begin{cases} 0 & \text{if } t \neq r \\ \sigma_\omega^2 \cdot Cor(h) & \text{if } t = r \end{cases} \quad \text{for } x \neq y$$

where $h = ||x - y|| \in \mathbb{R}$ and $Cor(h)$ is the Matérn correlation function.

The variance-covariance matrix can be written as:

$$\mathcal{C}(\delta(x, t), \delta(y, r)) = \frac{\varphi^{|t-r|}}{1 - \varphi^2} \sigma_\omega^2 Cor(h)$$

with $|t - r|$ the time lag between time step t and r.

Moving to spatio-temporal

The ideas are similar, but we add temporal correlations in the expression of the random effect δ.

For instance, let's introduce the model:

$$\delta(x, t) = \varphi \cdot \delta(x, t-1) + \omega(x, t) \text{ for } t = 2, ..., T$$

- $\varphi \in\,]-1; 1[$ is the autoregressive temporal term
- $\omega(x, t)$ is a purely spatial GRF
- $\omega(x, 1)$ derives from the stationary distribution $\mathcal{N}(0, \sigma^2/(1-\varphi^2))$

$$\mathcal{C}\left(\omega\left(x, t\right), \omega\left(y, r\right)\right) = \begin{cases} 0 & \text{if } t \neq r \\ \sigma_\omega^2 \cdot Cor(h) & \text{if } t = r \end{cases} \quad \text{for } x \neq y$$

where $h = ||x - y|| \in \mathbb{R}$ and $Cor(h)$ is the Matérn correlation function.

The variance-covariance matrix can be written as:

$$\mathcal{C}\left(\delta\left(x, t\right), \delta\left(y, r\right)\right) = \frac{\varphi^{|t-r|}}{1 - \varphi^2} \sigma_\omega^2 Cor(h)$$

with $|t - r|$ the time lag between time step t and r.

Such kind of process is:

- stationary: $\mathcal{C}(x, y; t, r) = \mathcal{C}(h, \tau)$ with $h = ||x - y||$ and $\tau = |t - r|$.
- separable: $\mathcal{C}(h, \tau) = \mathcal{C}^{(x)}(h) \cdot \mathcal{C}^{(t)}(\tau)$
- fully symmetric: $cov(\delta(x, t); \delta(y, r)) = cov(\delta(y, t); \delta(x, r))$

➠ But other non-separable covariance function exist and might be desirable to model complex interactions between spatial and temporal correlations.
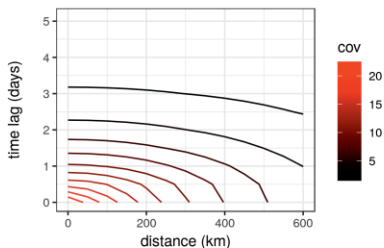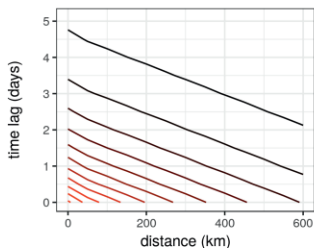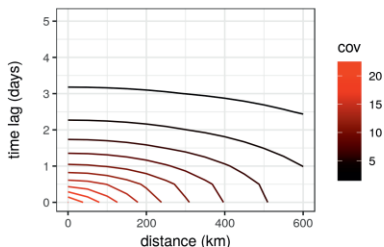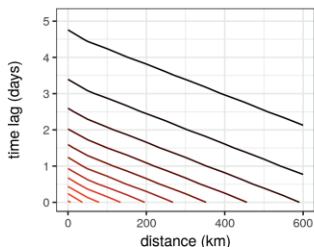

Separable covariance function (left) vs. non-separable covariance function (right)

Such kind of process is:

- stationary: $\mathcal{C}(x, y; t, r) = \mathcal{C}(h, \tau)$ with $h = ||x - y||$ and $\tau = |t - r|$.
- separable: $\mathcal{C}(h, \tau) = \mathcal{C}^{(x)}(h) \cdot \mathcal{C}^{(t)}(\tau)$
- fully symmetric: $cov(\delta(x, t); \delta(y, r)) = cov(\delta(y, t); \delta(x, r))$

➠ But other non-separable covariance function exist and might be desirable to model complex interactions between spatial and temporal correlations.

Separable covariance function (left) vs. non-separable covariance function (right)

Section 3

Inference method

Kriging

Basis of geostatistics ➠ Minimum mean-squared error approach

Let's define observations $\mathbf{Y} = (Y_1, Y_2, ..., Y_n)$

A linear predictor of Y_0 based on $\mathbf{Y}$ would take the form $\sum \ell_i Y_i + \alpha_0$

Using squared error loss, the best linear prediction would minimize

$$E\left[Y_0 - \left(\sum \ell_i Y_i + \alpha_0\right)\right]^2 \text{ over } \ell_i \text{ and } \alpha_0$$

Setting $\sum \ell_i = 1$, $\alpha_0 = 0$, $a_0 = 1$ and $a_i = -\ell_i$, the criterion becomes $E[\sum_{i=0}^n a_i Y_i]^2$ with $\sum a_i = 0$.

$$\text{And } E[\sum_{i=0}^n a_i Y_i]^2 = -\sum_i \sum_j a_i a_j \gamma(||x_i - x_j||)$$

The ℓ_i are then estimated by solving constrained optimization on this criterion

Kriging

Basis of geostatistics ➠ Minimum mean-squared error approach

Let's define observations $\mathbf{Y} = (Y_1, Y_2, ..., Y_n)$

A linear predictor of Y_0 based on $\mathbf{Y}$ would take the form $\sum \ell_i Y_i + \alpha_0$

Using squared error loss, the best linear prediction would minimize

$$E\left[Y_0 - \left(\sum \ell_i Y_i + \alpha_0\right)\right]^2 \text{ over } \ell_i \text{ and } \alpha_0$$

Setting $\sum \ell_i = 1$, $\alpha_0 = 0$, $a_0 = 1$ and $a_i = -\ell_i$, the criterion becomes $E[\sum_{i=0}^{n} a_i Y_i]^2$ with $\sum a_i = 0$.

$$\text{And } E[\sum_{i=0}^{n} a_i Y_i]^2 = -\sum_i \sum_j a_i a_j \gamma(||x_i - x_j||)$$

The ℓ_i are then estimated by solving constrained optimization on this criterion

Kriging

Basis of geostatistics ➡ Minimum mean-squared error approach

Let's define observations $\mathbf{Y} = (Y_1, Y_2, ..., Y_n)$

A linear predictor of Y_0 based on $\mathbf{Y}$ would take the form $\sum \ell_i Y_i + \alpha_0$

Using squared error loss, the best linear prediction would minimize

$$E\left[Y_0 - \left(\sum \ell_i Y_i + \alpha_0\right)\right]^2 \text{ over } \ell_i \text{ and } \alpha_0$$

Setting $\sum \ell_i = 1$, $\alpha_0 = 0$, $a_0 = 1$ and $a_i = -\ell_i$, the criterion becomes $E[\sum_{i=0}^{n} a_i Y_i]^2$ with $\sum a_i = 0$.

And $E[\sum_{i=0}^{n} a_i Y_i]^2 = -\sum_i \sum_j a_i a_j \gamma(||x_i - x_j||)$

The ℓ_i are then estimated by solving constrained optimization on this criterion

Kriging

Basis of geostatistics ➡ Minimum mean-squared error approach

Let's define observations $\mathbf{Y} = (Y_1, Y_2, ..., Y_n)$

A linear predictor of Y_0 based on $\mathbf{Y}$ would take the form $\sum \ell_i Y_i + \alpha_0$

Using squared error loss, the best linear prediction would minimize

$$E\left[Y_0 - \left(\sum \ell_i Y_i + \alpha_0\right)\right]^2 \text{ over } \ell_i \text{ and } \alpha_0$$

Setting $\sum \ell_i = 1$, $\alpha_0 = 0$, $a_0 = 1$ and $a_i = -\ell_i$, the criterion becomes $E[\sum_{i=0}^{n} a_i Y_i]^2$ with $\sum a_i = 0$.

$$\text{And } E[\sum_{i=0}^{n} a_i Y_i]^2 = -\sum_i \sum_j a_i a_j \gamma(||x_i - x_j||)$$

The ℓ_i are then estimated by solving constrained optimization on this criterion

Kriging

Basis of geostatistics ➠ Minimum mean-squared error approach

Let's define observations $\mathbf{Y} = (Y_1, Y_2, ..., Y_n)$

A linear predictor of Y_0 based on $\mathbf{Y}$ would take the form $\sum \ell_i Y_i + \alpha_0$

Using squared error loss, the best linear prediction would minimize

$$E\left[Y_0 - \left(\sum \ell_i Y_i + \alpha_0\right)\right]^2 \text{ over } \ell_i \text{ and } \alpha_0$$

Setting $\sum \ell_i = 1$, $\alpha_0 = 0$, $a_0 = 1$ and $a_i = -\ell_i$, the criterion becomes $E[\sum_{i=0}^{n} a_i Y_i]^2$ with $\sum a_i = 0$.

$$\text{And } E[\sum_{i=0}^{n} a_i Y_i]^2 = -\sum_i \sum_j a_i a_j \gamma(||x_i - x_j||)$$

The ℓ_i are then estimated by solving constrained optimization on this criterion

Maximum likelihood approach

Maximum likelihood methods consist in looking for the parameter values that maximize the likelihood

$$P(\mathbf{Y}|\boldsymbol{\theta})$$

For hierarchical models, this suppose to make an integral over the random effect $\boldsymbol{\delta}$ (with q the size of $\boldsymbol{\delta}$).

$$P(\mathbf{Y}|\boldsymbol{\theta}) = \int_{\mathbb{R}^q} P(\mathbf{Y}, \boldsymbol{\delta}|\boldsymbol{\theta}) d\delta$$

⟹ q is often very high in spatio-temporal model

To bypass the computational burden:
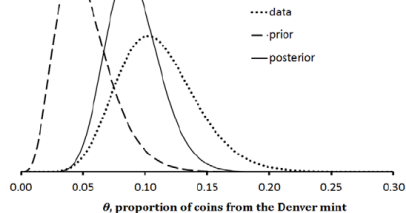- Laplace approximation
- Pseudo-likelihood

...

Maximum likelihood approach

Maximum likelihood methods consist in looking for the parameter values that maximize the likelihood

$$P(\mathbf{Y}|\boldsymbol{\theta})$$

For hierarchical models, this suppose to make an integral over the random effect $\boldsymbol{\delta}$ (with q the size of $\boldsymbol{\delta}$).

$$P(\mathbf{Y}|\boldsymbol{\theta}) = \int_{\mathbb{R}^q} P(\mathbf{Y}, \boldsymbol{\delta}|\boldsymbol{\theta}) d\delta$$

➠ q is often very high in spatio-temporal model

To bypass the computational burden:
- Laplace approximation
- Pseudo-likelihood

...

Maximum likelihood approach

Maximum likelihood methods consist in looking for the parameter values that maximize the likelihood

$$P(\mathbf{Y}|\boldsymbol{\theta})$$

For hierarchical models, this suppose to make an integral over the random effect $\boldsymbol{\delta}$ (with q the size of $\boldsymbol{\delta}$).

$$P(\mathbf{Y}|\boldsymbol{\theta}) = \int_{\mathbb{R}^q} P(\mathbf{Y}, \boldsymbol{\delta}|\boldsymbol{\theta}) d\delta$$

➡ q is often very high in spatio-temporal model

To bypass the computational burden:
- Laplace approximation
- Pseudo-likelihood

...

Bayesian inference

Bayesian Paradigm:

$$P(\boldsymbol{\theta}, \delta \mid \mathbf{Y}) = \frac{P(\mathbf{Y} \mid \boldsymbol{\theta}, \delta).P(\boldsymbol{\theta}, \delta)}{P(\mathbf{Y})}$$

- $P(\boldsymbol{\theta}, \delta \mid \mathbf{Y})$: posterior distribution of parameters and random effect

- $P(\mathbf{Y} \mid \boldsymbol{\theta}, \delta)$: likelihood of the data

- $P(\boldsymbol{\theta}, \delta)$: prior distribution of parameters and random effect

- $P(\mathbf{Y})$: marginal likelihood of the data (= normalizing constant)

MCMC: Random sampling in the joint posterior distribution $P(\delta, \theta | \mathbf{Y})$ (Metropolis-Hastings algorithm)

$$P(\delta, \theta | \mathbf{Y}) \propto P(\theta) \cdot P(\delta | \theta) \cdot P(\mathbf{Y} | \delta, \theta)$$
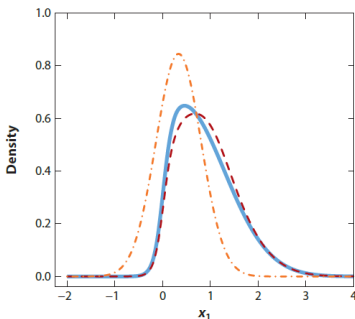
➜ **In practice, the estimation can be very long due to the size of δ**

⇒ **INLA:** instead of estimating the joint distribution $P(\delta, \theta | \mathbf{Y})$, approximate the marginal distribution $P(\theta | \mathbf{Y})$ and $P(\delta_i | \theta, \mathbf{Y})$.

$$\tilde{P}(\theta | \mathbf{y}) \propto \frac{P(\delta, \theta, \mathbf{Y})}{\tilde{P}_{\mathsf{G}}(\delta | \theta, \mathbf{Y})} \bigg|_{\delta = \delta^*(\theta)}$$

$$\tilde{P}(\theta_j | \mathbf{Y}) = \int \tilde{P}(\theta | \mathbf{Y}) d\theta_{-j}$$

$$\tilde{P}(\delta_i | \mathbf{y}) = \int \tilde{P}(\delta_i | \theta, \mathbf{Y}) \tilde{P}(\theta | \mathbf{Y}) d\theta$$


Marginal distribution of δ_1 (blue), Laplace approximation (red) and Gaussian approximation (orange).

MCMC: Random sampling in the joint posterior distribution $P(\delta, \theta | \mathbf{Y})$ (Metropolis-Hastings algorithm)

$$P(\delta, \theta | \mathbf{Y}) \propto P(\theta) \cdot P(\delta | \theta) \cdot P(\mathbf{Y} | \delta, \theta)$$

➜ **In practice, the estimation can be very long due to the size of δ**

➥ **INLA:** instead of estimating the joint distribution $P(\delta, \theta | \mathbf{Y})$, approximate the marginal distribution $P(\theta | \mathbf{Y})$ and $P(\delta_i | \theta, \mathbf{Y})$.

$$\tilde{P}(\theta | \mathbf{y}) \propto \left. \frac{P(\delta, \theta, \mathbf{Y})}{\tilde{P}_{\mathbf{G}}(\delta | \theta, \mathbf{Y})} \right|_{\delta = \delta^*(\theta)}$$

$$\tilde{P}(\theta_j | \mathbf{Y}) = \int \tilde{P}(\theta | \mathbf{Y}) d\theta_{-j}$$

$$\tilde{P}(\delta_i | \mathbf{y}) = \int \tilde{P}(\delta_i | \theta, \mathbf{Y}) \tilde{P}(\theta | \mathbf{Y}) d\theta$$

Marginal distribution of δ_1 (blue), Laplace approximation (red) and Gaussian approximation (orange).

Section 4

Some practice!

Data
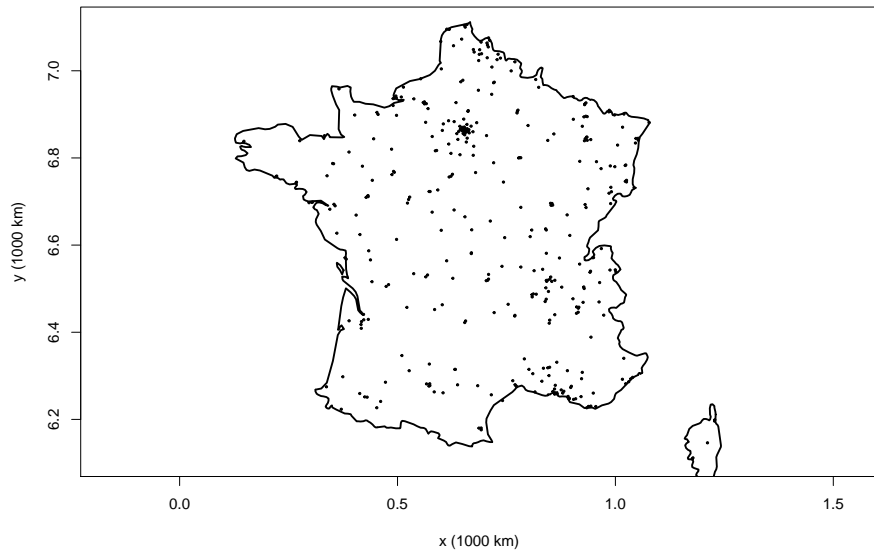
Creating the mesh
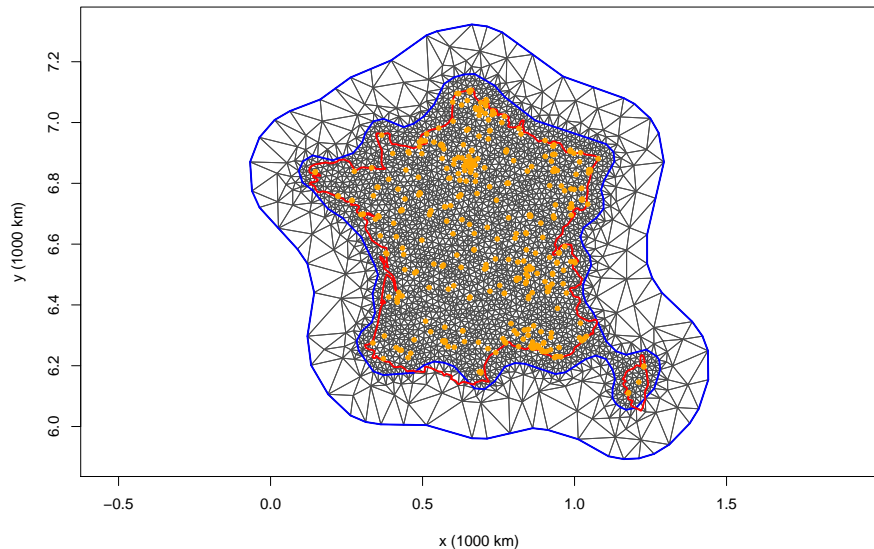
Building the SPDE model

```r
myspde=inla.spde2.pcmatern(mesh=mesh,
                           alpha=2,
                           prior.range=c(.01,.1),
                           prior.sigma=c(25,0.5))

# Observation matrix to make the connection
# between mesh nodes and observation sites:
A=inla.spde.make.A(mesh,loc=xy_OBS)

# Vector of indices for the nodes of the mesh
idx.spatial=inla.spde.make.index("spatial",
                                 n.spde=mesh$n)
```

Building the regression part

First, we create covariate dataframe: we add the intercept term 1, since in R-INLA it's more comfortable to include the intercept explicitly.

```
covar.df=data.frame(intercept=1,
                    x=xy_OBS[,1],
                    y=xy_OBS[,2])

# Create the stack with all the data and indices:
mystack=inla.stack(data=list(pm10=OBS_daily2014$PM10),
                   A=list(A,1),
                   effects=list(idx.spatial,covar.df))

# We write the formula for the model:
myformula=pm10~-1+intercept+f(spatial,model=myspde)
```

Note: since we handle the intercept explicitly, we have to include "-1"). Recall that we have given the name "spatial" to the index of the spatial effect.

Fitting the model with R-INLA

The following `inla` run should take approximatively 1 minute.

```
fit=inla(myformula,
         data=inla.stack.data(mystack),
         family="gaussian",
         control.predictor=list(A=inla.stack.A(mystack),
                                 compute=FALSE),
         control.inla=list(int.strategy="eb",
                           strategy="gaussian"),
         verbose=F)
```

```
##
## Call:
##    c("inla(formula = myformula, family = \"gaussian\", data =
##    inla.stack.data(mystack), ", " verbose = F, control.predictor = list(A
##    = inla.stack.A(mystack), ", " compute = FALSE), control.inla =
##    list(int.strategy = \"eb\", ", " strategy = \"gaussian\"))")
## Time used:
##     Pre = 2.12, Running = 106, Post = 0.506, Total = 109
## Fixed effects:
##             mean    sd 0.025quant 0.5quant 0.975quant   mode kld
## intercept 18.039 0.323     17.404   18.039     18.674 18.039   0
##
## Random effects:
##    Name    Model
##      spatial SPDE2 model
##
## Model hyperparameters:
##                                          mean    sd 0.025quant 0.5quant
## Precision for the Gaussian observations 0.009 0.000      0.009    0.009
## Range for spatial                       0.022 0.005      0.013    0.022
## Stdev for spatial                       4.248 0.358      3.659    4.205
##                                         0.975quant  mode
## Precision for the Gaussian observations      0.009 0.009
## Range for spatial                            0.032 0.022
## Stdev for spatial                            5.054 4.085
##
## Expected number of effective parameters(stdev): 196.69(0.00)
## Number of equivalent replicates : 381.75
##
## Marginal log-Likelihood:  -283189.38
```

Predictions with R-INLA

To build predictions, the input data (`inla.stack`) must be slightly modified. NA values are added to the vector of observations and inla will predict latent field values at these locations.

```r
# Observation matrix to make the connection between mesh nodes and observation sites:
proj_grid=inla.mesh.projector(mesh,
                              xlim=range(bound$loc[,1]),
                              ylim=range(bound$loc[,2]),
                              dims=c(100,100))

# Grid on wich we calculate the predictions:
xygrid=as.matrix(expand.grid(proj_grid$x,proj_grid$y))

# Put prediction coordinates and observation coordinates into A:
A=inla.spde.make.A(mesh,
                   loc=rbind(xygrid,xy_OBS))

# We add the intercept term 1, since in RINLA it's more comfortable to include the intercept explicitly.
covar.df=data.frame(intercept=1,
                    x=c(xygrid[,1],
                        xy_OBS[,1]),
                    y=c(xygrid[,2],
                        xy_OBS[,2]))

# Create the stack with all the data and indices: add NA data here
mystack=inla.stack(data=list(pm10=c(rep(NA,nrow(xygrid)),
                                    OBS_daily2014$PM10)),
                   A=list(A,1),
                   effects=list(idx.spatial,covar.df),
                   tag="mytag")
```

When fitting, we now set `compute=TRUE` (to calculate fitted and predicted values) and `link=1` (although only necessary when using non-identity link).
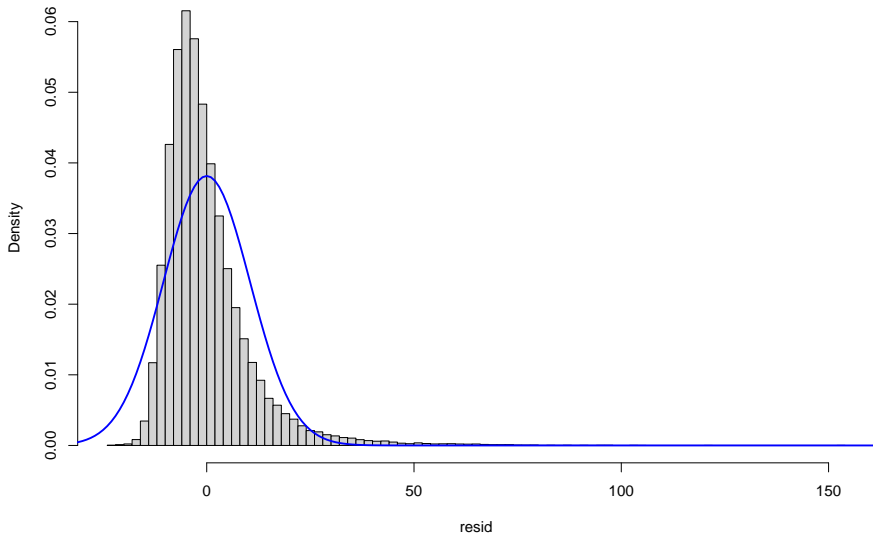
Running INLA can now take considerably longer since we calculate the posterior estimation `y.hat` for each observation point y (`compute=TRUE`).

```
fit=inla(myformula,
         data=inla.stack.data(mystack),
         family="gaussian",
         control.predictor=list(A=inla.stack.A(mystack),
                                 compute=TRUE,
                                 link=1),
         control.inla=list(int.strategy="eb",
                           strategy="gaussian"),
         verbose=TRUE)
```
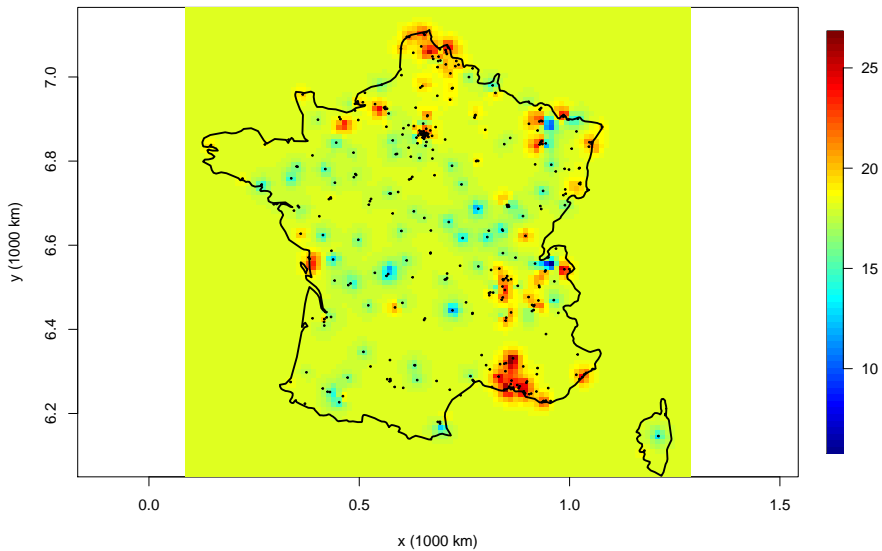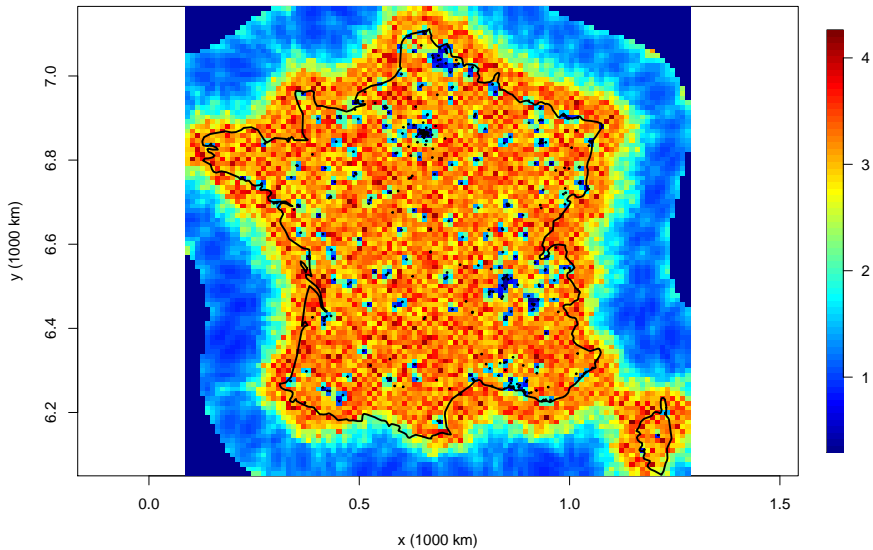
Histogram of resid

Spatio-temporal application
Simulating and preparing the data
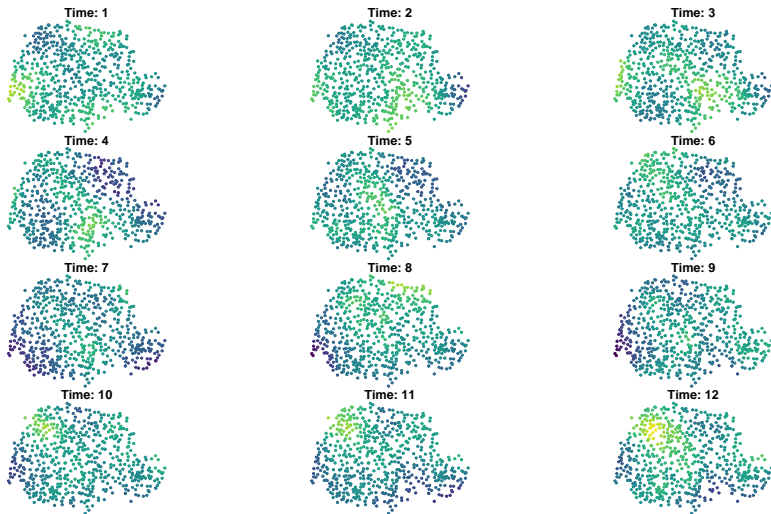
Fitting the model - model estimates
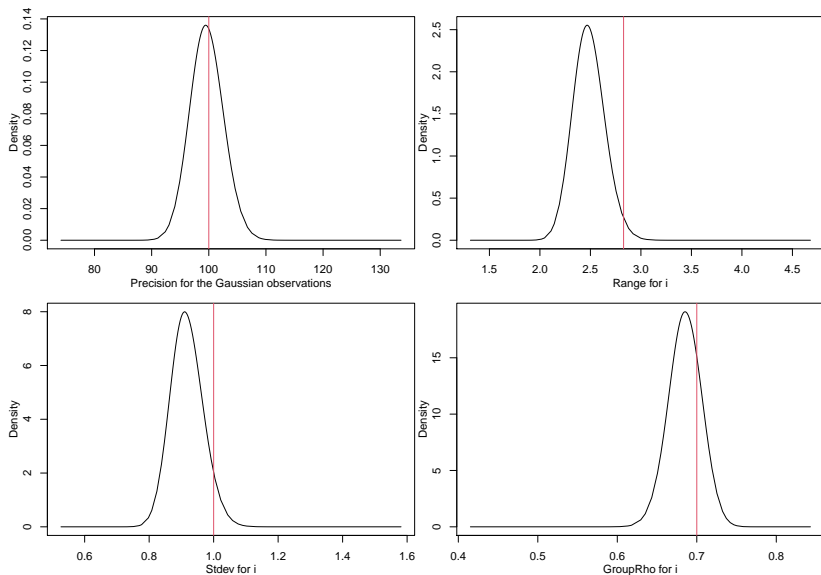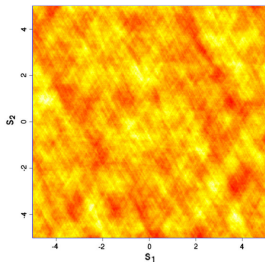
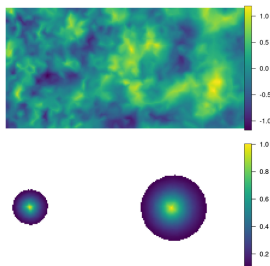# Fitting the model - model spatial predictions
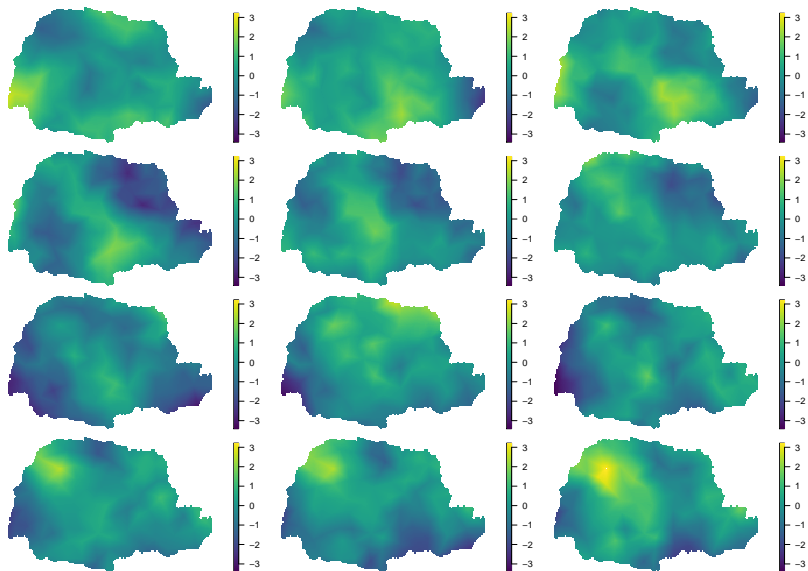
Section 5

Few words on some subject that we did not address here

Few words on some subject that we did not address here
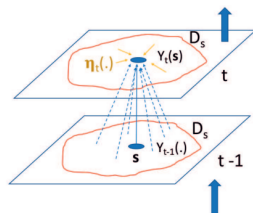
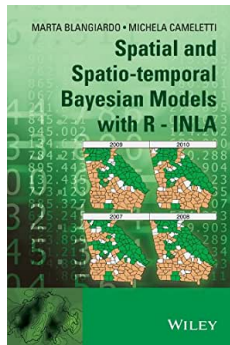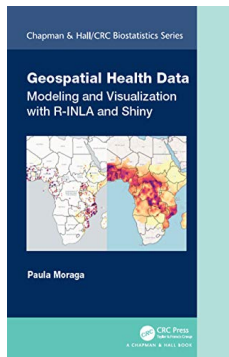**Anisotropy** **Non-stationarity** **Dynamic modeling**

Section 6

Some useful references

Some useful references
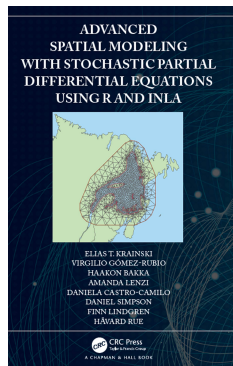
And **Spatio-Temporal Statistics with R** by Christopher K. Wikle, Andrew Zammit-Mangion, and Noel Cressie. Published by Chapman & Hall/CRC. https://spacetimewithr.org/

Section 7

Work in autonomy

Work in autonomy

- Explore some aspects of spatial statistics that I did not presented here (anisotropy, non-stationarity, EOF, aggregation indices, etc)

- Apply INLA on available dataset (EVHOE or anything else)

- Present some nice modeling/coding stuff based on previous references

➠ 2 hours of work in autonomy and restitution