Spatio-temporal geostatistics

ECAS2019: Statistical Analysis for Space-Time Data

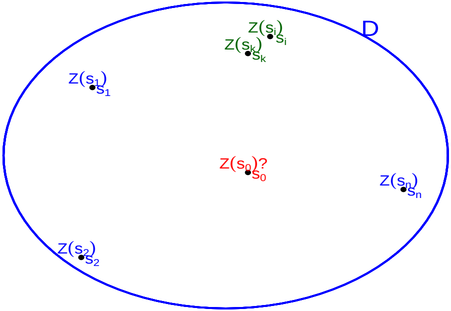
Lisboa, Portugal

15-17 July, 2019

A brief history of geostatistics

- '60 Foundation at MinesParisTech by G. Matheron
- '90 Mostly mining and petroleum applications; some environment
  Very little space $\times$ time
- '10 most data are spatio-temporal

Georges Matheron, Wikipedia



$D \subset \mathbb{R}^2$
continuous

$Z = \{Z(s)\}_{s \in D}$ is a stochastic process (random field) i.e.,
a set of random variables indexed by $s \in D$; $Z(s_1), Z(s_2), \ldots, Z(s_n)$ observed,

Questions

- Estimate dependence relationships (covariance function) of Z?
- Predict (maping) $Z(s_0)$, $Z(s_0)$ not observed or $\int_B Z(s)ds$?
- Simulations, conditional simulations ?
- Spatial modeling (regression with spatial dependence) ?

Why is it important to take into account spatial dependence ?

- Spatial dependence induces redundancy in the data
- The equivalent number of independent data is lower (because the correlation is generally positive)
- Not taking into account the spatial correlation leads to
  - errors in confidence intervals
  - errors in p-values, and hence in test conclusions
- Spatial dependence brings information : when the locations are nearby the associated random variables are correlated (Tobler's law) $\hookrightarrow$ useful for prediction

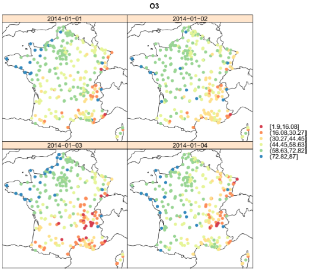
Spatio temporal data

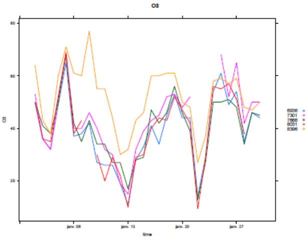
- Data are collected with space (S) and time (T) coordinates
- We wish to account for space-time interaction
  - trends
  - dependencies

$\hookrightarrow$ Generalization of both spatial statistics and time series

Several space-time data configuration are possible

Dense in time, scarce in space

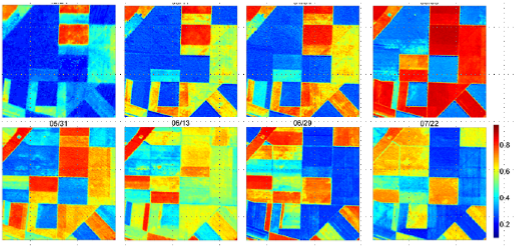




- Use multivariate time series?
- What model in space? How do we interpolate at a new site?

Spatio temporal data

Dense in space, scarce in time



- Use multivariate spatial statistics
- What evolution in time ? How do we perform temporal prediction ?
- One possibility : spatial field $\otimes$ ARMA process

Spatio temporal data

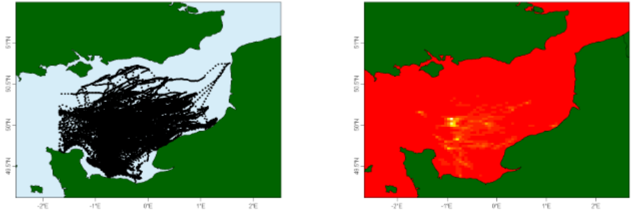
Other configurations

Dense in space and time

- No particular emphasis on space and/or time
- special interest in the space × interaction
- Focus on prediction at unmeasured locations

Even more complex

- no time alignment
- scarcity in space and time
- data along trajectories from moving devices, ...



(a) Observations *(b) Agrégation*

P. Gloaguen, PhD

Spatio temporal data

Some challenges

- Relatively large datasets
  $N_T \times N_S \times N_p$ is usually very large. "Big N problem" in all aspects of data manipulation: memory, vizualization, computations, ...

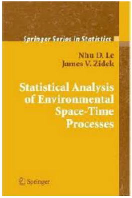
- Statistical model
  space-time interaction: time arrow, causality and physical laws
  $\hookrightarrow$ specific covariance models

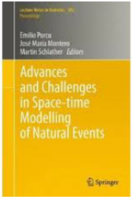
- Estimation (fitting) and Prediction (kriging) $\hookrightarrow$ max likelihood usually intractable; specific algorithms are required

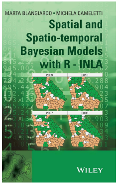
- Simulation: few algorithms; still lot to do

Spatio temporal data

Some books







a paper

Journal de la Société Française de Statistique
Vol. 158 No. 3 (2017)

Analyzing spatio-temporal data with R: Everything you always wanted to know – but were afraid to ask

Titre: Donnees spatio-temporelles avec R : tout ce que vous avez toujours voulu savoir sans jamais avoir osé le demander

RESSTE Network et al.[1,2]

https://informatique-mia.inra.fr/resste/paper-workshop

French pollution data

- Ambient air pollution assessed through monitoring network
  - 4 pollutants : ozone, nitrogen dioxide, particulate matter PM_{10}, PM_{25}
  - daily and hourly data for the year 2014
  - 507 stations (rural, suburban,urban)
  - 42% missing data
- Air quality modeling simulating physical and chimical processses
  - chemistry-transport model *CHIMERE*
  - hourly time step
  - rectangular grid covering France resolution 10km

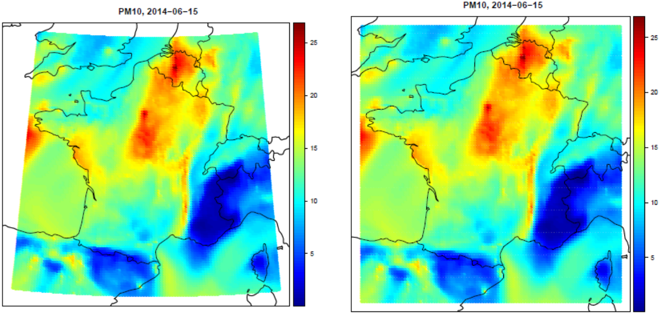


FIGURE 1. *Lambert93 (left) vs WSG84 (right) coordinates reference systems for the French* PM_{10} *concentration forecasts by the* CHIMERE *model on the* 15^{th} *of June 2014.*

Covariance models

Statistical model

$$Z(\mathbf{s},t) = \mu(\mathbf{s},t) + Y(\mathbf{s},t) + \varepsilon(\mathbf{s},t) \qquad (\mathbf{s},t) \in D \times T \subset \mathbb{R}^d \times \mathbb{R}$$

- $\mu(\mathbf{s},t)$ is a trend term

$$\mu(\mathbf{s},t) = \sum_{k=1}^{p} \beta_k f_k(X_k(\mathbf{s})) + \sum_{\ell=1}^{q} \alpha_\ell f_\ell(X_\ell(t))$$

- $Y(\mathbf{s},t)$ is a spatially correlated random effect
- $\varepsilon(\mathbf{s},t)$ is an independent error (nugget effect)

Stationary process

$Y(\mathbf{s},t)$ is a second order stationary random field i. e.

$$\begin{aligned} \mathrm{E}(Y(\mathbf{s},t)) &= \mu(= 0) \\ \mathrm{Cov}(Y(\mathbf{s}+\mathbf{h},t+u), Y(\mathbf{s},t)) &= C(\mathbf{h},u) \end{aligned}$$

Covariance functions

Spatial random field

Property

If $C(\cdot)$ is the covariance function of a second order stationary process, then

- $\forall h \in \mathbb{R}^2 \quad C(h) = C(-h)$
- $C(0) = \mathit{Var}(Z(s))$
- $\forall h \in \mathbb{R}^2 \quad |C(h)| \leq C(0)$
- $C(\cdot)$ *is positive definite i.e.*
  $\forall s_1, \ldots, s_n, \quad \forall \lambda_1, \ldots, \lambda_n \quad \sum_{ij} \lambda_i \lambda_j C(s_i - s_j) \geq 0$

Property

- The sum and the product of two positive definite functions are positive definite functions.
- A linear combination with positive coefficients is a positive definite function.

Stationary increments

Definition

Z is a process with stationary increment if the increments of Z are second-order stationary, i.e.

$$E(Z(s) - Z(s')) = 0 \qquad Var(Z(s) - Z(s')) = 2\gamma(s - s')$$

Property

- $\gamma(h) \geq 0$, $\gamma(0) = 0$, $\gamma(-h) = \gamma(h)$
- $\lim_{\|h\| \to \infty} \dfrac{\gamma(h)}{\|h\|^2} = 0$
- *if the random process is second order stationary*

$$\gamma(h) = C(0) - C(h)$$

- *if* $\lim_{|h| \to \infty} \gamma(h) = \ell < +\infty$ *then the random process is second order stationary and* $\ell = C(0)$

Variogram

Definition

An authorized linear combination (ALC) for a random process Z is a linear combination such that:

$$Var\big(\sum_i \lambda_i Z(s_i)\big) < +\infty$$

Proposition

The authorized linear combination (ALC) for a random process Z with stationary increments are such that

$$\sum_i \lambda_i = 0$$

Property

The variogram function is conditionally negative definite i.e.

$$\forall s_1, \ldots, s_n, \quad \forall \lambda_1, \ldots, \lambda_n \textit{ with } \textstyle\sum_{i=1}^n \lambda_i = 0$$
$$\textstyle\sum_{ij} \lambda_i \lambda_j \gamma(s_i - s_j) \leq 0$$

Variogram

Some admissible variogram functions

- constant $\gamma(h) = C$
- exponential $\gamma(h) = C(1 - \exp(-|h|/\rho)) \quad C > 0 \quad \rho > 0$
- spherical $\gamma(h) = \begin{cases} C\big(\frac{3}{2}\frac{|h|}{\rho} - \frac{1}{2}\frac{|h|^3}{\rho^3}\big) & \text{if} \quad 0 \leq |h| \leq \rho \\ C & \text{if} \quad |h| \geq \rho \end{cases}$
- Gaussian $\gamma(h) = C(1 - \exp(-|h|^2/\rho)) \quad C > 0 \quad \rho > 0$
- power $\gamma(h) = C|h|^\alpha \quad \alpha < 2$
- ...

Matern class

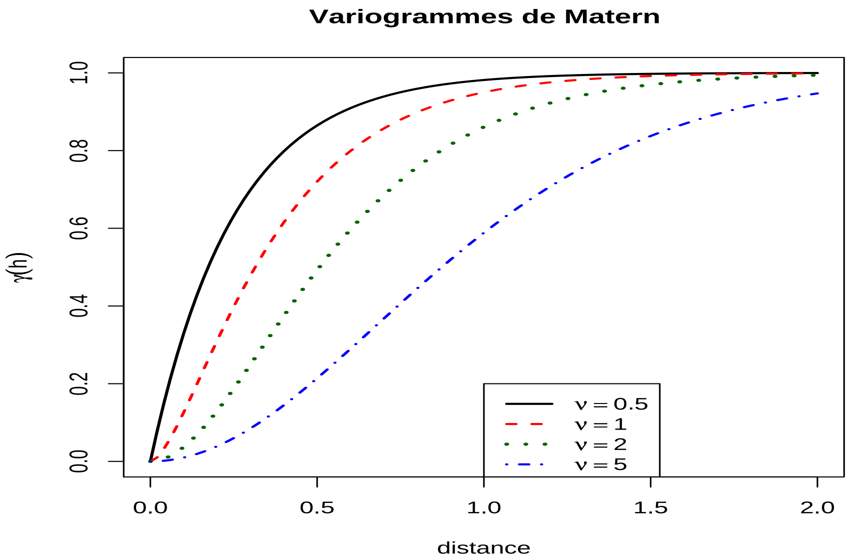
$$\gamma(h) = C\Big[1 - \frac{1}{2^{\nu-1}\Gamma(\nu)}\Big(\frac{2\nu^{1/2}h}{\rho}\Big)\mathcal{K}_\nu\Big(\frac{2\nu^{1/2}h}{\rho}\Big)\Big]$$

$\mathcal{K}_\nu$: modified Bessel function of the third kind , order ν

ν : tunes the regularity near 0.

- $\nu = 1/2$: exponential
- $\nu \to +\infty$: Gaussian

Variogram



Variogrammes de Matern
γ(h)
0.0
0.2
0.4
0.6
0.8
1.0
0.0
0.5
1.0
1.5
2.0
distance
ν = 0.5
ν = 1
ν = 2
ν = 5

Separability

Definition

The spatio-temporal covariance function C_{ST} is separable if

$$C_{ST}(\mathbf{h}, u) \propto C_S(\mathbf{h}) C_T(u) \qquad \forall \mathbf{h}, u$$

It is equivalent to conditional independence:

$$Z(\mathbf{s}, t) \perp Z(\mathbf{s}', t') | Z(\mathbf{s}, t') \qquad \forall \mathbf{s}, \mathbf{s}' \in D; \quad \forall t, t' \in T$$

Advantages

- Easy to understand; less memory and easier coding
- Matrix computation is fast thanks to the Kronecker product

Drawbacks

- No complex space-time interaction
- not realistic in many applications, especially environmental.

$\hookrightarrow$ Towards non separable covariance models

Non-separability

Definition

The spatio-temporal covariance function C_{ST} is fully symmetric if

$$C_{ST}(\mathbf{h}, u) = C_{ST}(-\mathbf{h}, u) = C_{ST}(\mathbf{h}, -u) = C_{ST}(-\mathbf{h}, -u) \qquad \forall \mathbf{h}, u$$

Separability $\Longrightarrow$ full symmetry

- Positively non separable if

$$\rho_{ST}(\mathbf{h}, u) > \rho_S(\mathbf{h})\rho_T(u)$$

- Negatively non separable if

$$\rho_{ST}(\mathbf{h}, u) < \rho_S(\mathbf{h}h)\rho_T(u)$$

- Some covariance functions are not uniformly positively or negatively non separable

Separability factor

$$C_{ST}(\mathbf{h}, u)/\sigma^2_{ST} - C_S(\mathbf{h})C_T(u)/\sigma^4_{ST}$$

is 0 when separable

Some non separable models

- Product sum model (De Iaco, Myers, Posa, 2001)

$$C(\mathbf{h}, u) = a_0 C_S^0(\mathbf{h}) C_T^0(u) + a_1 C_S^1(\mathbf{h}) + a_2 C_T^2(u) \qquad a_0, a_1, a_2 > 0$$

  Negatively non separable

- "Frozen model" (Briggs, 1968)

$$C(\mathbf{h}, u) = C_S(\mathbf{h} - vu)$$

  where v is a velocity

- Gneiting class

$$C(\mathbf{h}, u) = \frac{\sigma^2}{\psi(u^2)^{1/2}} \varphi\left(\frac{\|\mathbf{h}\|}{\psi(u^2)}\right) \qquad (\mathbf{h}, u) \in \mathbb{R}^d \times \mathbb{R}$$

  where $\varphi(t)$ is a completely monotonic function and $\psi(t)$ is a positive function such that $\psi'(t)$ is completely monotonic.
  Positively non separable

An ubiquitous model

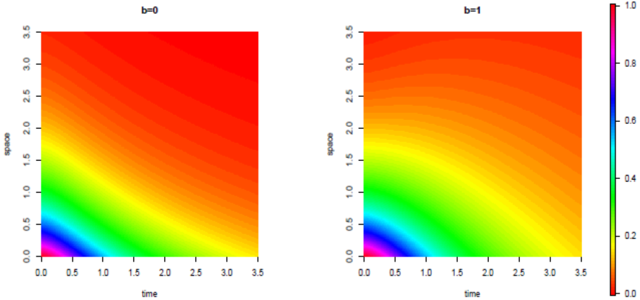
Gneiting-Matern class of spatio-temporal covariance model

$$\varphi(\mathbf{h}) = \mathcal{M}(\mathbf{h}; r, \nu) \qquad \psi(t) = (at^a + 1)^b, \qquad t \geq 0$$

with $\alpha, r, \nu > 0$, $0 < a \leq 1$, $0 \leq b \leq 1$

$$C(\mathbf{h}, u) = \frac{\sigma^2}{(\alpha|u|^{2a} + 1)^{b/2}} \mathcal{M}\Big(\frac{\mathbf{h}}{(\alpha|u|^{2a} + 1)^b}; r, \nu\Big) \qquad (\mathbf{h}, u) \in \mathbb{R}^d \times \mathbb{R}$$

Can be generalized to multivariate data (Bourotte, Allard and Porcu, 2017)

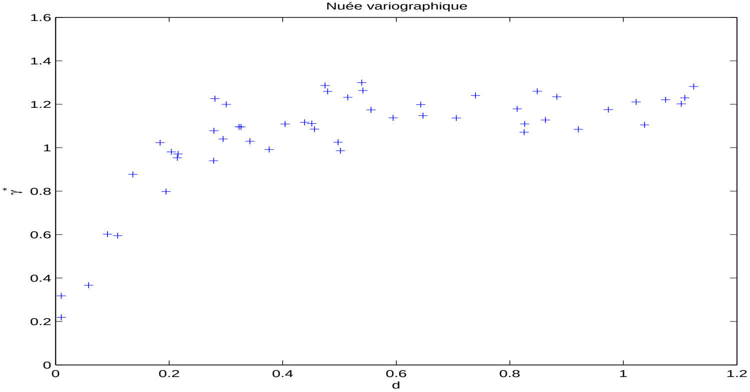


Variogram estimation

Spatial random field

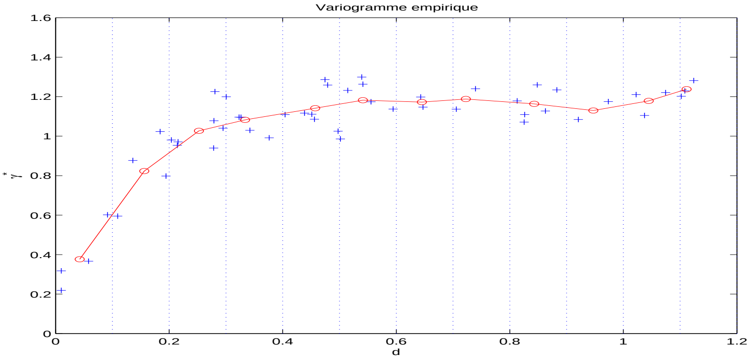
Variogram cloud

$$\gamma^*_{ij} = \frac{(z(s_i) - z(s_j))^2}{2} \qquad d_{ij} = \|s_i - s_j\|$$



Empirical variogram

$$\widetilde{\gamma}(d_k) = \frac{1}{2n_c} \sum_{i,j \in C(k)} (Z(s_i) - Z(s_j))^2 \quad i,j \in C(k) \Leftrightarrow (k-1)\delta \leq \|s_i - s_j\| \leq k\delta$$

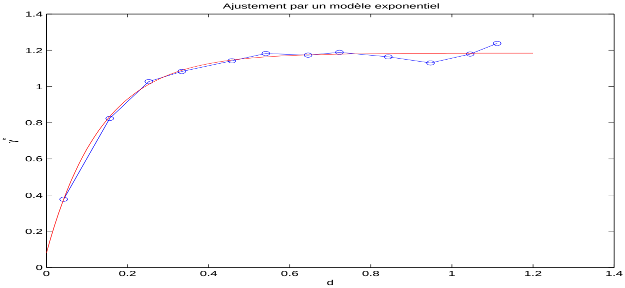


Variogram estimation

Fitting the variogram

$\gamma_{\theta,\tau}$ an admissible variogram, $(\theta\tau)$ are solution of

$$\min_{\theta,\tau} \sum_{k=1}^{K} (\gamma_{\theta,\tau}(d_k) - \widetilde{\gamma}(d_k))^2$$



Estimating the covariance function

Parameters $\theta = (\mu, \sigma^2, r_S, r_T, b, \ldots)$

- Fitting: Empirical variogram + (weighted) least squares (`gstat` , `Rgeostats` , `RandomFields`)
- Maximum Likelihood: (`RandomFields`)
- Tapering: multiply $C(\mathbf{h}, u)$ with a compactly supported covariance function $\hookrightarrow$ sparse structure.
- Composite likelihood, e.g. Pairwise Likelihood (`CompRandFld`)

Pairwise Likelihood (PL)

Definition

$$PL(\theta, z) = \prod_{i,j} f_{\theta;i,j}(z_i, z_j)$$

$f_{\theta;i,j}(z_i, z_j)$ is one of the following density

$$(z(\mathbf{s}_i, t_i), z(\mathbf{s}_j, t_j)) \qquad z(\mathbf{s}_i, t_i)|z(\mathbf{s}_j, t_j) \qquad z(\mathbf{s}_i, t_i) - z(\mathbf{s}_j, t_j)$$

- PL estimator

$$\widehat{\theta}_{PL} = \mathrm{argmax}_{\theta}\mathrm{PL}(\theta; \mathrm{z})$$

- Reduced computations if restricted to pairs that are close in space and time
- Good statistical properties (consistence, bias) but $\mathrm{Var}(\widehat{\theta}_{\mathrm{PL}}) \leq \mathrm{Var}(\widehat{\theta}_{\ell})$
- Not efficient for confidence intervals, hypothesis testing

Model selection

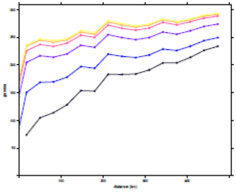
Choose the type of separability (negative, null, positive)

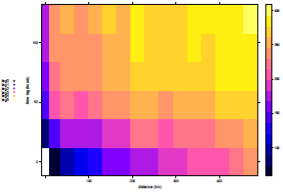
- Visual comparison (very difficult in a spatio-temporal context, even more so if multivariate)
- CLIC (Composite Likelihood Information Criterion)
- Cross validation (find the correct scheme)

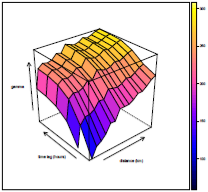
Estimating correctly the spatio-temporal dependence is of paramount importance.

Empirical spatio-temporal variogram

```
plot(vvst)                    # sample variograms at each time lag
plot(vvst,map=FALSE)          # ST-variogram map
plot(vvst,wireframe=TRUE) # ST-variogram wireframe
```







Fitting the variogram

Variogram models with gstat

```
## separable model, least squares fit
separableModel <- vgmST("separable",space=vgm(0.9,"Exp",10000,0.1),
                        time=vgm(0.9,"Exp",3.5,0.1),sill=40)

## product sum model, least squares fit
ProductSum <- vgmST("productSum",space =vgm(9,"Exp",8e3,1),
                     time=vgm(8,"Exp",106,2),k=2)

## metric model, "manual" fit
Metric <- vgmST("metric",joint=vgm(60,"Exp",8e3,10,
                add.to=vgm(70,"Exp",1e5,0,stAni=1000)),stAni=150)

## sum-metric model, manual fit
SumMetric <- vgmST("sumMetric",space=vgm(30,"Exp",20000),time=vgm(40,"Exp",
                   joint=vgm(50,"Exp",1.6e5,0),stAni=1000)
```

Empirical spatio-temporal variogram

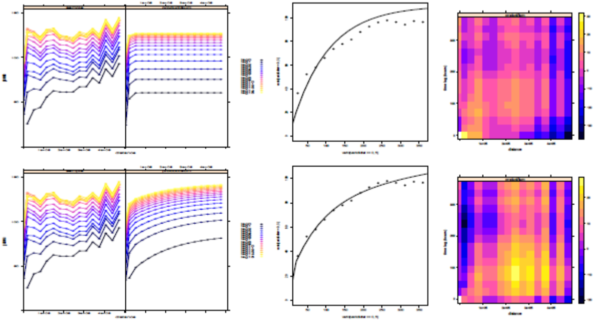


FIGURE 6. *Empirical and fitted variograms for a product-sum model. Upper line: WLS fit with* `fit.StVariogram` *Lower line: manual fit. Left: empirical vs. fitted variogram for different time lags (*`plot` *arguments* `all=T,map=F`*). Middle: marginal temporal variogram. Right: difference map of space-time variogram values (*`plot` *arguments* `diff=T,map=T`*).*

Fitting the variogram

Variogram models with `CompRandFld`

```
#space-time separable exponential model without nugget
cormod1 <- "exp_exp"
fixed1  <- list(mean=mean_est,nugget=0,sill=var_est)
start1  <- list(scale_s=200,scale_t=2)
#space-time separable exponential model with nugget
cormod2 <- "exp_exp"
fixed2  <- list(mean=mean_est,sill=var_est)
start2  <- list(scale_s=200,scale_t=2,nugget=0)
#Gneiting model with powers fixed to 1, without nugget
cormod3 <- "gneiting"
fixed3  <- list(sill=var_est,mean=mean_est,nugget=0,power_s=1,power_t=1)
start3  <- list(scale_s=200,scale_t=2,sep=.5)
#Gneiting model with powers fixed to 1, with nugget
cormod4 <- "gneiting"
fixed4  <- list(sill=var_est,mean=mean_est,power_s=1,power_t=1)
start4  <- list(scale_s=200,scale_t=2,sep=.5)
```

Fitting the variogram

```
#Gneiting model with powers fixed to .5, without nugget
cormod5 <- "gneiting"
fixed5  <- list(sill=var_est,mean=mean_est,nugget=0,power_s=.5,power_t=.5)
start5  <- list(scale_s=200,scale_t=2,sep=.5)
#Gneiting model with powers fixed to .5, with nugget
cormod6 <- "gneiting"
fixed6  <- list(sill=var_est,mean=mean_est,power_s=.5,power_t=.5)
start6  <- list(scale_s=200,scale_t=2,sep=.5)
#Gneiting model where powers are estimated, without nugget
#difficult to propose good starting values, but we will take estimated paran
cormod7 <- "gneiting"
fixed7  <- list(sill=var_est,mean=mean_est,nugget=0)
#Gneiting model where powers are estimated, with nugget
#difficult to propose good starting values, but we will take estimated paran
cormod8 <- "gneiting"
fixed8  <- list(sill=var_est,mean=mean_est)
```

Fitting spatio-temporal variogram

Empirical spatio-temporal variogram

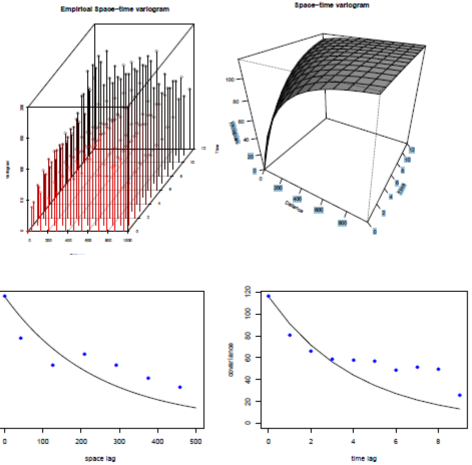


FIGURE 5. *Visual diagnostics for model 1 adjusted with the composite likelihood approach of* `CompRandFld`*. Above: Empirical and fitted space-time variograms, here plotted with the* `Covariogram` *function. Below: Empirical and fitted marginal covariance functions for space (left) and time (right).*

Spatial prediction: kriging

$Z(s_1) = z(s_1), \ldots, Z(s_n) = z(s_n)$ observed, predict $Z(s_0)$

Optimal predictor:

$$p^{opt} = \mathsf{E}\big(Z(s_0) | Z(s_1) = z(s_1), \ldots, Z(s_n) = z(s_n)\big)$$

minimize quadratic risk

$$\mathsf{E}\big((Z(s_0) - \widehat{Z}(s_0))^2\big)$$

Best Linear Unbiased Predictor(BLUP) :

$$p^* = \mu + \sum_{i=1}^{n} \lambda_i Z(s_i)$$

If Z Gaussian, $p^{opt} = p^*$

Solution

$$\begin{aligned} \lambda &= C^{-1}c \qquad C_{i,j} = \mathsf{cov}\big(Z(s_i), Z(s_j)\big) \quad c_i = \mathsf{cov}\big(Z(s_0), Z(s_i)\big) \\ \mu &= \mathsf{E}(Z(s_0)) - \sum_{i=1}^{n} \lambda_i \mathsf{E}(Z(s_i)) \end{aligned}$$

Simple kriging

D. Krige : South African mining engineer

C and $m(s)$ known

$$\widehat{Z(s_0)} = {}^t c C^{-1}(Z - m) + m(s_0)$$

Kriging variance

$$\sigma_{SK}^2 = \mathsf{E}\big((Z(s_0) - \widehat{Z}(s_0))^2\big) = \sigma^2(s_0) -{}^t cC^{-1}c$$

If $s_0 = s_i$, then

$$\widehat{Z(s_i)} = Z(s_i)$$

C et $m(s)$ have to be estimated.

Ordinary kriging

$Z(s) = \mu + \delta(s)$, with stationary increments.

$\widehat{Z}(s_0) = \sum_{i=1}^{n} \lambda_i Z(s_i)$ such that

$$E(\widehat{Z}(s_0)) = E(Z(s_0)) \text{ and } E(\widehat{Z}(s_0) - Z(s_0))^2 \text{ minimum}$$

variogram : $\gamma(h) = \frac{1}{2}\mathsf{Var}(Z(s+h) - Z(s))$

$(\lambda_i)_{i=1,n}$, $(\sum_{i=1}^{n} \lambda_i = 1)$ solution of the linear system

$$\begin{pmatrix} 0 & \gamma(s_1 - s_2) & \dots & \gamma(s_1 - s_n) & 1 \\ \gamma(s_1 - s_2) & 0 & \dots & \gamma(s_2 - s_n) & 1 \\ \dots & \dots & \dots & \dots & \dots \\ \gamma(s_1 - s_n) & \gamma(s_1 - s_n) & \dots & 0 & 1 \\ 1 & 1 & \dots & 1 & 0 \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \dots \\ \lambda_n \\ \alpha \end{pmatrix} = \begin{pmatrix} \gamma(s_0 - s_1) \\ \gamma(s_0 - s_2) \\ \dots \\ \gamma(s_0 - s_n) \\ 1 \end{pmatrix}$$

Ordinary kriging

Kriging variance

$$\sigma_K^2(s_0) = E((\widehat{Z}(s_0) - Z(s_0))^2) = \alpha + \sum_{i=1}^{n} \lambda_i \gamma(s_i - s_0)$$

Remarks

i) The λ_i may be negative, or greater than 1.
ii) If $s_0 \in \{s_1, \dots s_n\}$ then $\lambda_i = 1$, $\lambda_j = 0$, $j \neq i$, and $\sigma_K^2(s_i) = 0$
iii) The kriging weights depend on the arrangement of the measurement locations, on the location of the prediction, on the sample size and on the variogram function.

Universal kriging

If there is a spatial trend

$$Z(s) = \mu(s) + \delta(s) \qquad \delta \text{ with stationary increments } \mathsf{E}(\delta) = 0$$

Modelling the trend $\mathsf{E}(Z(s)) = \mu(s)$, $\mu(s) = \sum_{j=0}^{p} \beta_j f_j(s)$

$f_0(s) = 1$, $f_1(s) = x$, $f_2(s) = y$, $f_3(s) = xy, \ldots$, $s = (x, y)$,
(β_j) unknown

Linear predictor $\widehat{Z}(s_0) = \sum_{i=1}^{n} \lambda_i Z(s_i)$

Unbiased conditions (universality):

$$\textstyle\sum_i \lambda_i f_k(s_i) = f_k(s_0),\ k = 0, \ldots p$$

Universal kriging

$\gamma(\cdot)$ is the δ-variogram

$$\begin{pmatrix} 0 & \dots & \gamma(s_1 - s_n) & f_0(s_1) & \dots & f_p(s_1) \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \gamma(s_1 - s_n) & \dots & 0 & f_0(s_n) & \dots & f_p(s_n) \\ f_0(s_1) & \dots & f_0(s_n) & & & \\ \dots & \dots & \dots & & 0 & \\ f_p(s_1) & \dots & f_p(s_n) & & & \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \\ \alpha_0 \\ \vdots \\ \alpha_p \end{pmatrix} = \begin{pmatrix} \gamma(s_0 - s_1) \\ \vdots \\ \gamma(s_0 - s_n) \\ f_0(s_0) \\ \vdots \\ f_p(s_0) \end{pmatrix}$$

Kriging variance $\sigma^2_{KU}(s_0) = \sum_{i=1}^{n} \lambda_i \gamma(s_0 - s_i) + \sum_{j=1}^{p+1} \alpha_{j-1} f_{j-1}(s_0)$

Universal kriging

δ-variogram estimation

$$\frac{1}{2}(\delta(s_i) - \delta(s_j))^2 = \frac{1}{2}\Big(Z(s_i) - Z(s_j) - \sum_{k=1}^{p+1} \beta_{k-1}(f_{k-1}(s_i) - f_{k-1}(s_j))\Big)^2$$

β_i have to be estimated.

$$\widehat{\beta} = (X^t \Sigma^{-1} X)^{-1} X^t \Sigma^{-1} Z \qquad \Sigma = \mathrm{cov}(\delta)$$

Iterative procedure:

- estimate β using OLS,
- fit the variogram from the residuals,
- estimate β using GLS,
- ...

Maximum likelihood : estimate regression parameters and variogram parameters, assuming $Z(s)$ Gaussian

Kriging with non constant mean

External drift :

$$Z(s) = \beta_0 + \beta_1 \varphi(s) + \delta(s)$$

$\varphi(s)$ known $\forall s \in D$

$$\widehat{Z}(s_0) = \sum_{i=1}^{n} \lambda_i Z(s_i)$$

Regression kriging :

$$Z(s) = \beta X(s) + \delta(s)$$

$$\widehat{\beta} = (X^t X)^{-1} X^t Z \quad , \quad \widehat{\delta}(s_i) = Z(s_i) - X(s_i)\widehat{\beta}$$

$$\widehat{Z}(s_0) = \widehat{\beta} X(s_0) + \sum_{i=1}^{n} \lambda_i \widehat{\delta}(s_i)$$

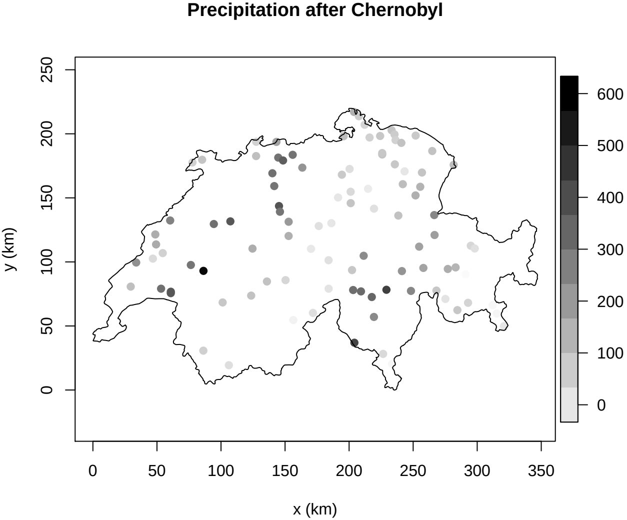
Conditional simulations

Objective: Simulate random fields according to a distribution such that the values at the measurement sites equal the observed values.

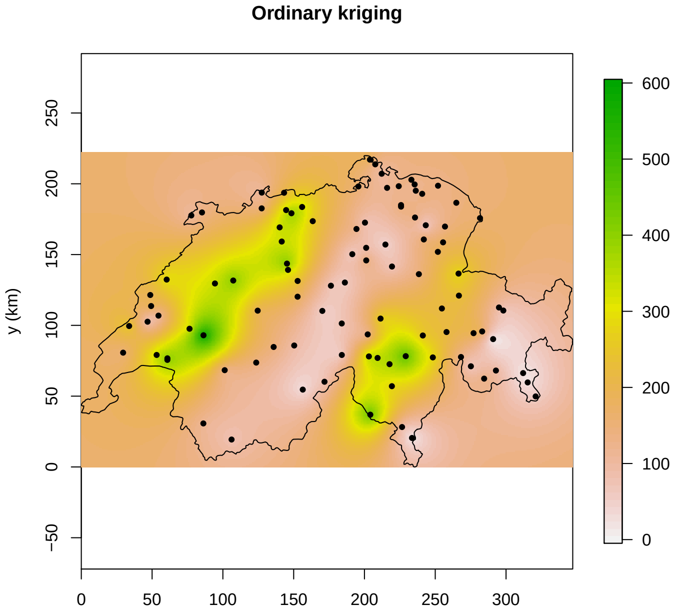
Conditional kriging

- estimate the covariance function $C(h)$
- $\widehat{Z} = \sum_i \lambda_i Z_i$ kriging observations
- simulate a random field Y on D with covariance function $C(h)$
- $\widehat{Y} = \sum_i \lambda_i Y_i$ kriging the simulated values at the measurement locations
- $\widehat{Z} + Y - \widehat{Y}$ is a conditional simulation.

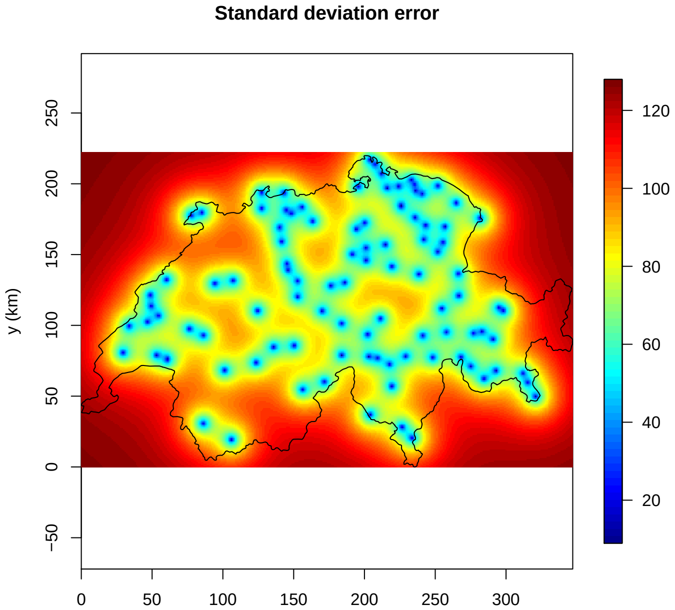
Chernobyl rainfall in Switzerland



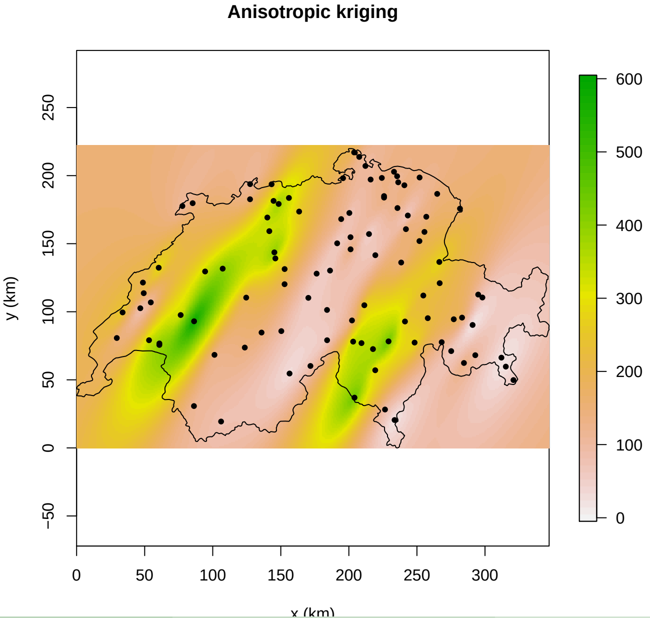
Chernobyl rainfall in Switzerland



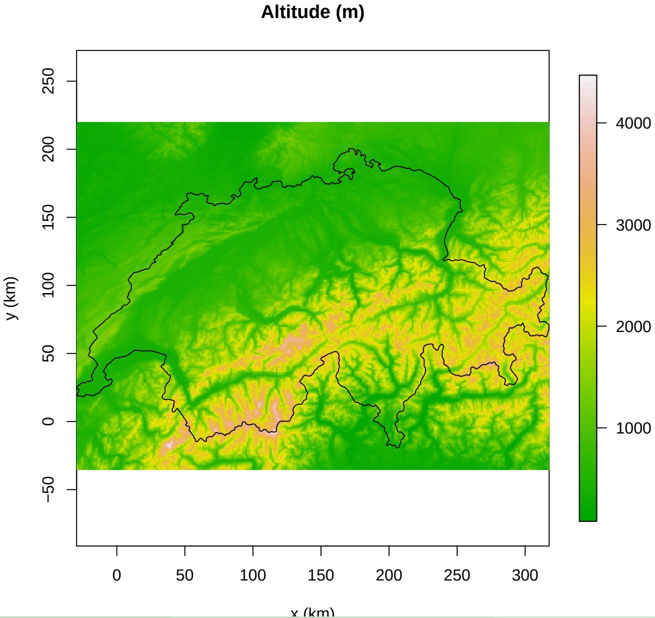
Chernobyl rainfall in Switzerland



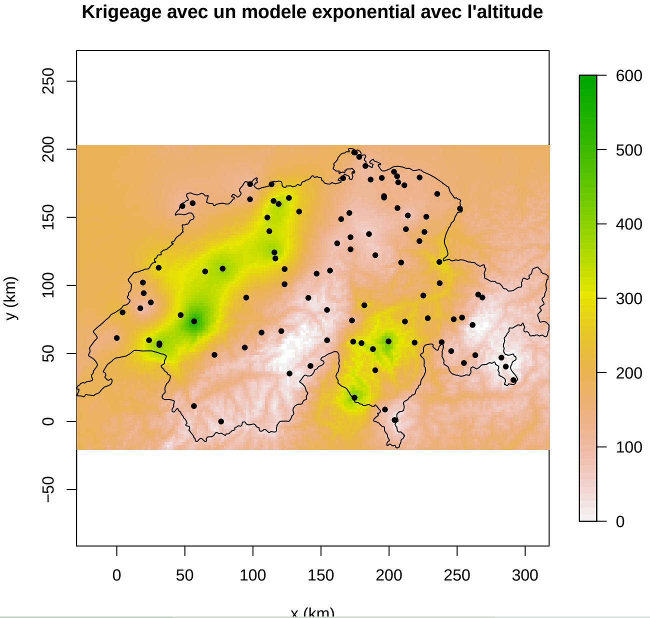
Chernobyl rainfall in Switzerland



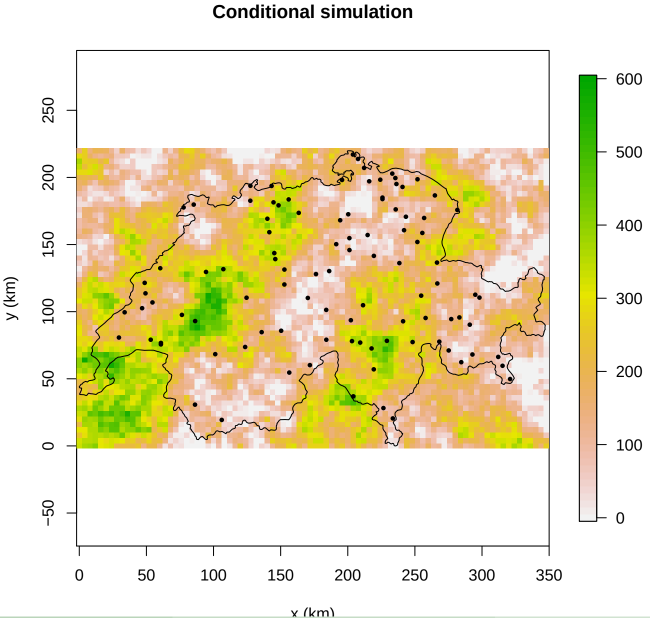
Chernobyl rainfall in Switzerland



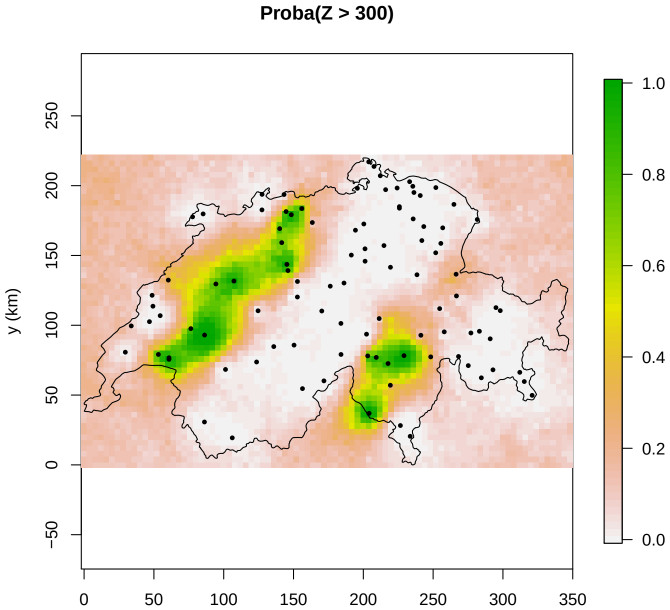
Chernobyl rainfall in Switzerland



Chernobyl rainfall in Switzerland



Chernobyl rainfall in Switzerland



Cokriging

2 random fields $Z_1(.), Z_2(.)$
$(s_{ik}, k = 1, n_i)$ measurement locations of field Z_i.

- isotopic case: $s_{1k} = s_{2k} = s_k$
- heterotopic case: locations do not coincide at all (total), or partially (partial).

Stationary assumptions:

- $\mathsf{E}\big(Z_i(s)\big) = \mu_i$, $\mathsf{cov}(Z_i(s), Z_j(s+h)) = c_{ij}(h)$
- if Z isotropic $c_{ij}(h) = c_{ij}(\|h\|)$

Properties

- $c_{ij}(h) = c_{ji}(-h)$
- $|c_{ij}(h)| \leq \sqrt{c_{ii}(0)c_{jj}(0)}$

Cokriging

Stationary joint increments

- $\mathsf{E}(Z_j(s+h) - Z_j(s)) = 0$,
- $\mathsf{cov}(Z_i(s+h) - Z_i(s), Z_j(s+h) - Z_j(s)) = 2\gamma_{ij}(h)$

Cross variogram

if stationary $\gamma_{ij}(h) = c_{ij}(0) - \frac{1}{2}(c_{ij}(h) + c_{ij}(-h))$

Cross pseudo-variogram

$$\begin{aligned} \widetilde{\gamma}_{ij}(h) &= \frac{1}{2}\mathsf{Var}(Z_i(s+h) - Z_j(s)) \\ &= \frac{1}{2}(c_{ii}(0) + c_{jj}(0)) - c_{ij}(-h) \text{ if stationary} \end{aligned}$$

Cokriging

Linear predictor

$$\widehat{Z}_1(s) - \mu_1(s) = \sum_{k=1}^{n_1} \lambda_{1k}(Z_1(s_{1k}) - \mu_1(s_{1j})) + \sum_{k=1}^{n_2} \lambda_{2k}(Z_2(s_{2k}) - \mu_2(s_{2j}))$$

Minimize $\mathsf{E}(\widehat{Z}_1(s) - Z_1(s))^2$ such that $\mathsf{E}(\widehat{Z}_1(s) - Z_1(s)) = 0$

$$\sum_{j=1}^{2}\sum_{l=1}^{n_j} \lambda_{jl}\gamma_{ij}(s_{ik} - s_{jl}) + \alpha_i \quad = \quad \gamma_{1i}(s_{ik} - s) \quad i = 1,2 \quad k = 1 \dots n_i$$

$$\sum_{k=1}^{n_1} \lambda_{1k} = 1 \qquad \sum_{k=1}^{n_2} \lambda_{2k} = 0$$

Cokriging variance

$$\sigma^2_{CO} = \sum_{i=1}^{2}\sum_{k=1}^{n_i} \lambda_{ik}\gamma_{1i}(s_{ik} - s) + \alpha_1$$

Cokriging

Variograms estimation

$\gamma_{ij}(h)$ conditionally negative definite.

Coregionalization models:

$$\gamma_{ij}(h) = \sum_{\ell=0}^{L} b_{ij}^{l} \gamma_\ell(h) \qquad \Gamma(h) = \sum_{l=0}^{L} B_\ell \gamma_\ell(h)$$

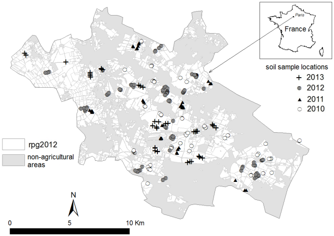
Each B_l must be positive definite, i.e.

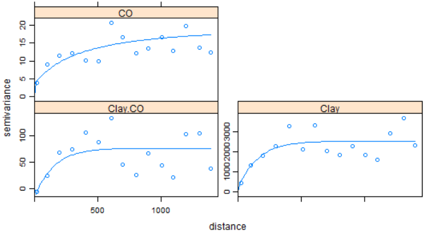
$$b_{11}^{\ell} \geq 0 \quad ; \quad b_{22}^{\ell} \geq 0 \quad ; \quad |b_{12}^{\ell}| \leq \sqrt{b_{11}^{\ell} b_{22}^{l}}$$

Cokriging : topsoil organic carbon

Map the topsoil organic carbon in an agricultural region, using measurements acquired from soil samples (SOC, pH, granulometry) and a DEM (Digital Elevation model)

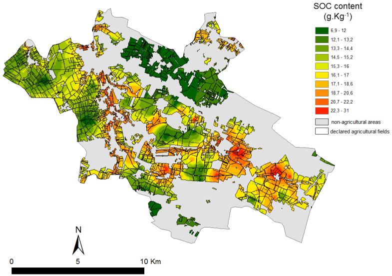
$$\widehat{C}(s_0) = \mu + \alpha Elev(s_0) + \sum_{i=1}^{n} \lambda_i C(s_i) + \sum_{k=1}^{K} \sum_{i=1}^{n} \beta_i^k X^k(s_i)$$



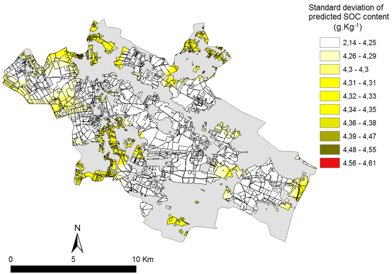


Package RGeostats

Cokriging : topsoil organic carbon



SOC prediction



cokriging standard deviation

Zaouche et al., 2017

Multivariate kriging

Z^1, Z^2,...,Z^p correlated random fields to predict
Cross-covariance function

$$C_{ij}(s, s') = \mathsf{cov}(Z^i(s), Z^j(s')) = C_{ij}(s - s')$$

Separable

$$C_{ij}(h) = a_{ij} C(h)$$

$A = [a_{ij}]_{i,j=1\ldots,p}$ positive definite matrix

Non separable (Gneiting)

$$C_{ij}(h) = \sigma_i \sigma_j r_{ij} C(h, \rho_{ij}, \nu_{ij})$$

$C(h, \rho, \nu)$ Matern correlation function, range ρ, order ν
$\rho_{ij} = (\rho_i^2 + \rho_j^2)/2$, $\nu_{ij} = (\nu_i + \nu_j)/2$, condition on r_{ij}

Spatio-temporal kriging

We suppose $C(\mathbf{h}, u)$ is known (i.e. use plug-in estimator)

Best Linear Unbiaised Predictor

$$Z^*(\mathbf{s}_0, t_0) = \sum_{\mathbf{s}_\alpha, k} \lambda_{\alpha,k} Z(\mathbf{s}_\alpha, t_k)$$

with $t_0 \neq t_k$ and/or $\mathbf{s}_0 \neq s_\alpha$

Many possible contexts

- Prediction : $t_0 \neq t_k$
- Cross validating a whole series at $\mathbf{s}_0$: $\mathbf{s}_0 \neq \mathbf{s}_\alpha$
- Analysis : $t_k \in \{t_0, t_{-1}, t_{-2}, \dots\}$ but $(\mathbf{s}_\alpha, t_k) \neq (\mathbf{s}_0, t_0)$
- Reanalysis: $t_k \in \{\dots, t_2, t_1, t_0, t_{-1}, t_{-2}, \dots\}$ but $(\mathbf{s}_\alpha, t_k) \neq (\mathbf{s}_0, t_0)$

Spatio-temporal kriging

Simple kriging

When the expectation is 0,

$$Z^*(s_0, t_0) = C_0^t C^{-1} Z$$

with $C = \mathrm{Var(Z)}$ and $C_0 = \mathrm{Cov(Z(s_0, t_0), Z)}$

When $n_S \times n_T$ is large , C^{-1} is impossible to compute. Need to use

- Moving neighborhood; slow; introduces discontinuities; somehow arbitrary; but good at handling non-stationarity
- "Tapering": $\Longrightarrow$ inversion uses sparse matrix computation
- "Low rank" models: C is replaced by a low rank approximation $\Longrightarrow$ over-smoothing
- INLA/SPDE: uses sparse precision matrix representation+sparse matrix computation. `R-INLA`.
  Very efficient in Space. Spatio-temporal models currently limited to separable AR(1) or RW models in time

Geostatistical regression model

$$Z(s,t) = \mu(s,t) + Y(s,t) + \varepsilon(s,t) \qquad (s,t) \in D \times T \subset \mathbb{R}^2 \times \mathbb{R}$$

with

$$\mu(s,t) = \sum_{k=1}^{p} \beta_k f_k(X_k(s)) + \sum_{\ell=1}^{q} \beta_\ell f_k(X_\ell(t))$$

Catch-22 situation

- Need to know $\mu(s,t)$ for estimating $C(h,u)$
- Need to know $C(h,u)$ for estimating $\mu(s,t)$

Simultaneaous estimation with ML usually intractable in spatio-temporal except with SPDE/INLA with AR(1) or RW in time

$\hookrightarrow$ use an iterative approach

Validation

External Validation (EV)

- Split the dataset in two parts, training set and validation set

Cross Validation (CV)

- Leave-one-out cross validation
- K-fold CV (less costly)

For EV and K-fold CV, one can imagine subsets adapted to the spatio-temporal framework

- spatial : the same validation set of stations for all times
- temporal : all the stations for a subset of times

Prediction scores

- Mean square error and Normalized mean square error

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(\mathrm{Z}_i^* - Z_i)^2 \text{ and } \mathrm{NMSE} = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{\mathrm{Z}_i^* - Z_i}{\sigma_i^*}\right)^2$$

- Logarithmic score

$$\mathrm{LogS} = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{2}\log(2\pi\sigma_i^{2*}) + \frac{1}{2}\left(\frac{\mathrm{Z}_i^* - Z_i}{\sigma_i^*}\right)^2\right)$$

- Continued Ranked Probability Score

$$\frac{1}{n}\sum_{i=1}^{n}\int_{-\infty}^{\infty}(F_i(z) - \mathbf{1}_{Z_i^* \leq z})^2 dy$$

with $F_i(z) = \mathbb{P}(Z_i \leq y | Z_j, j \neq i)$

R implementation

- CompRandFld

  ```
  Kri(...)
  ```

  enables using tapering covariance
  non conditional simulations via `RFism`
- gstat

  ```
  krigeST(...)
  ```

  neighborhood may be defined by number max of neighbors, unique in the standard case
  no simulations
- RandomFields

  ```
  RFinterpolate(...)
  ```

  neighborhood defined by number max of neighbors
  conditional and non-conditional simulations via `RFismulate`

CompRandFld

```
Kri(loc,time,coordx,coordt,corrmodel, param, data)
```

with

- `data` : input data, matrix format
- `loc` : locations to predict coordinates, matrix format
- `time` : times to predict
- `coordx` : data coordinates
- `coordt` : data times
- `corrmodel` : correlation model
- `param` : correlation parameters

Returns a matrix of size $nrow(coordx) \times length(coordt)$.

```
krigeST(formula, data, newdata, modelList, beta,
         nmax = Inf, computeVar = FALSE)
```

with

- `formula` : for modelling the trend with possible covariates. ex `PM10~1` for ordinary kriging
- `data` : data, ST format
- `newdata` : locations to predict coordinates, ST format
- `modelList` : variogram model, vgmST format
- `beta` : known mean for simple kriging
- `nmax` : number max of neighbors for kriging with sliding neighborhood. Unique neighborhood standard case.
- `computeVar` : kriging variance computation
- other options useless

Returns an object class ST with the predicted values

RandomField

```
RFinterpolate(model, x, y = NULL, z = NULL, T = NULL,
grid=NULL, distances, dim, data, given=NULL,
err.model, ignore.trend = FALSE, ...)
```

with

- `model` covariance model with parameters
- `data` input data, array format (x_i, y_i, z_i, t_i, regular in t_i) or RFsp
- `x`, `y`, `z` locations to predict coordinates, regular in t
- `grid` locations to predict on a grid

Returns a vector of length $nx \times ny \times nt$

PM_{10} concentration data

$$Z(\mathbf{s},t) = \mu(\mathbf{s},t) + Y(\mathbf{s},t) + \varepsilon(s\mathbf{s},t)$$

$\mu(\mathbf{s},t)$: output of CHIMERE model $\hookrightarrow$ kriging residuals

- 103 stations without missing data
- 3 days of data at all observation locations
- locations to predict : the $10km \times 10km$ CHIMERE grid
- predict at four different times : t_{-2}, t_0, t_1, t_2

TABLE 4. *Leave-one-out RMSE and one-day-ahead $RMSE_{ext}$ computed for 6 estimated spatio-temporal covariance models by leave-one-out cross-validation and external validation.. Sep. = Separable model; ProdSum = Product Sum model. Autofit is the automatic fitting using `fit.STVariogram` Manual indicates manual fit by visual inspection. Metric and SumMetric can only be fitted manually*

Model	Sep. Autofit	Sep. Manual	ProdSum Autofit	ProdSum Manual	Metric	SumMetric
RMSE	9.73	9.90	9.62	10.03	9.87	10.43
$RMSE_{EXT}$	12.74	14.35	12.74	13.79	14.70	13.43

TABLE 3. *Leave-one-out RMSE and one-day-ahead $RMSE_{EXT}$ computed for 8 estimated spatio-temporal covariance models by leave-one-out cross-validation and external validation.*

Model	1	2	3	4	5	6	7	8
RMSE	10.18	9.68	9.78	9.53	9.71	9.74	9.77	9.75
$RMSE_{EXT}$	12.85	12.08	12.32	11.97	12.57	12.51	12.51	12.36

PM$_{10}$ concentration data

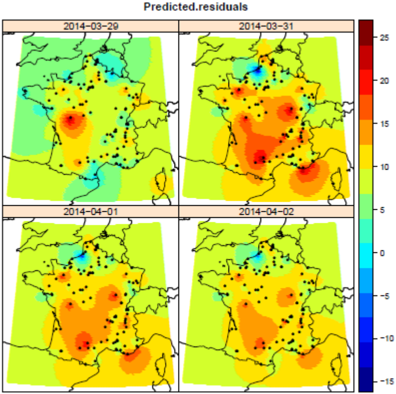


FIGURE 7. *Predicted residuals over France with CompRandFld.*

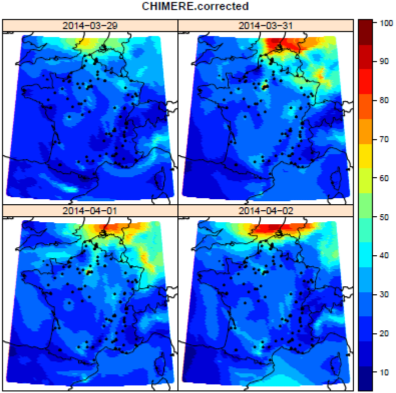


FIGURE 10. *Corrected PM$_{10}$ over France with CompRandFld.*

Miscellaneous

- Another point of view : functional kriging (Nerini 2010, Giraldo 2013)

$$\chi : t \in \mathbb{R} \rightarrow \chi(t) \in \mathbb{R}$$

$$\widehat{\chi}(s_0) = \sum_{i=1}^{n} \lambda_i \chi(s)$$

  functional variogram or covariance

- Kriging on a sphere : specific covariance functions (Porcu 2018)
- Kriging compositional data (Allard 2017)