Geostatistical simulations

N. Desassis
From C. Lantuejoul's course

MINES Paris - Geosciences

ATHENS Geostatistics 2024

Geostatistics in a probabilistic framework

- Aims:
  - Predict if a variable is above a cut-off
  - Improve the spatial predictor
  - Have a better measure of the prediction error
  - Make simulations

Geostatistics in a probabilistic framework

- Aims:
  - Predict if a variable is above a cut-off
  - Improve the spatial predictor
  - Have a better measure of the prediction error
  - Make simulations

Price to pay : add some modeling assumptions

Geostatistics in a probabilistic framework

- Aims:
  - Predict if a variable is above a cut-off
  - Improve the spatial predictor
  - Have a better measure of the prediction error
  - Make simulations

Price to pay : add some modeling assumptions

Modeling the spatial law :

$$P(Z(x_1) \leq z_1, \ldots, Z(x_n) \leq z_n)$$

$$\forall n \in \mathbb{N}^\star, \forall (x_1, \ldots, x_n) \in (\mathbb{R}^d)^n, \forall (z_1, \ldots, z_n) \in \mathbb{R}^n$$

Geostatistics in a probabilistic framework

- Aims:
  - Predict if a variable is above a cut-off
  - Improve the spatial predictor
  - Have a better measure of the prediction error
  - Make simulations

  Price to pay : add some modeling assumptions

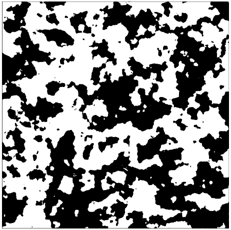
  Modeling the spatial law :

$$P(Z(x_1) \leq z_1, \ldots, Z(x_n) \leq z_n)$$

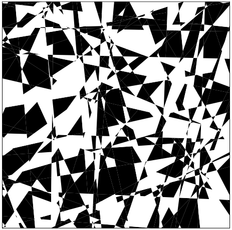
$$\forall n \in \mathbb{N}^\star, \forall (x_1, \ldots, x_n) \in (\mathbb{R}^d)^n, \forall (z_1, \ldots, z_n) \in \mathbb{R}^n$$

- Outline
  - Gaussian model (convenient)
  - Gaussian transform (to honor the "marginal" distribution)

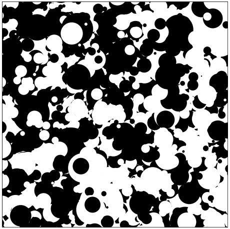
Spatial law for complete characterisation



Gaussian excursion set



Poisson polygons



Dead leaves

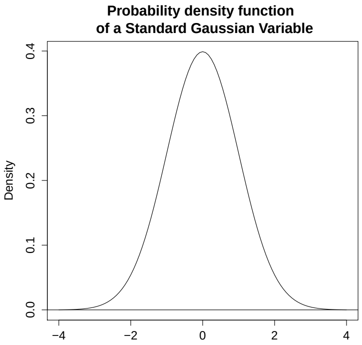
Same marginal distributions (Bernouilli), same variogram (same bivariate distributions) and even same trivariate distributions

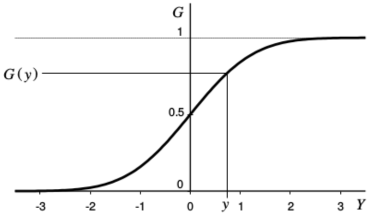
Marginal distributions and second-order moments (covariance) are not sufficient to characterize the processus!

Gaussian random variable

A random variable Y is Gaussian with mean m and variance σ^2 ($Y \sim \mathcal{N}(m, \sigma^2)$) if it has the following probability density function (p.d.f)

$$g_{m,\sigma^2}(y) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(y-m)^2}{2\sigma^2}\right), \qquad y \in \mathbb{R}$$





Second-order stationary Gaussian Random Function

Mean m and covariance function C

A random function $(Y(x), x \in \mathbb{R}^d)$ is second-order stationary and Gaussian if

Second-order stationary Gaussian Random Function

Mean m and covariance function C

A random function $(Y(x), x \in \mathbb{R}^d)$ is second-order stationary and Gaussian if

$$\forall n \in \mathbb{N}^\star,$$

Second-order stationary Gaussian Random Function

Mean m and covariance function C

A random function $(Y(x), x \in \mathbb{R}^d)$ is second-order stationary and Gaussian if

$$\forall n \in \mathbb{N}^\star, \forall (x_1, \ldots, x_n) \in (\mathbb{R}^d)^n,$$

Second-order stationary Gaussian Random Function

Mean m and covariance function C

A random function $(Y(x), x \in \mathbb{R}^d)$ is second-order stationary and Gaussian if

$$\forall n \in \mathbb{N}^\star, \forall (x_1, \ldots, x_n) \in (\mathbb{R}^d)^n,$$

$$\mathbf{Y} = (Y(x_1), \ldots, Y(x_n)) \text{ is a Gaussian vector}$$

Second-order stationary Gaussian Random Function

Mean m and covariance function C

A random function $(Y(x), x \in \mathbb{R}^d)$ is second-order stationary and Gaussian if

$$\forall n \in \mathbb{N}^\star, \forall (x_1, \ldots, x_n) \in (\mathbb{R}^d)^n,$$

$$\mathbf{Y} = (Y(x_1), \ldots, Y(x_n)) \text{ is a Gaussian vector}$$

with

Second-order stationary Gaussian Random Function

Mean m and covariance function C

A random function $(Y(x), x \in \mathbb{R}^d)$ is second-order stationary and Gaussian if

$$\forall n \in \mathbb{N}^\star, \forall (x_1, \ldots, x_n) \in (\mathbb{R}^d)^n,$$

$$\mathbf{Y} = (Y(x_1), \ldots, Y(x_n)) \text{ is a Gaussian vector}$$

with

- $E[Y(x_i)] = m$

Second-order stationary Gaussian Random Function

Mean m and covariance function C

A random function $(Y(x), x \in \mathbb{R}^d)$ is second-order stationary and Gaussian if

$$\forall n \in \mathbb{N}^\star, \forall (x_1, \ldots, x_n) \in (\mathbb{R}^d)^n,$$

$$\mathbf{Y} = (Y(x_1), \ldots, Y(x_n)) \text{ is a Gaussian vector}$$

with

- $E[Y(x_i)] = m$
- The covariance matrix of Y is $\Sigma = (\sigma_{ij})$ with

$$\sigma_{ij} = C(x_i - x_j)$$

Second-order stationary Gaussian Random Function

Mean m and covariance function C

A random function $(Y(x), x \in \mathbb{R}^d)$ is second-order stationary and Gaussian if

$$\forall n \in \mathbb{N}^\star, \forall (x_1, \ldots, x_n) \in (\mathbb{R}^d)^n,$$

$$\mathbf{Y} = (Y(x_1), \ldots, Y(x_n)) \text{ is a Gaussian vector}$$

with

- $E[Y(x_i)] = m$
- The covariance matrix of Y is $\Sigma = (\sigma_{ij})$ with

$$\sigma_{ij} = C(x_i - x_j)$$

- The spatial distribution is entirely characterized by the expectation m and the covariance function C.

Conditional distribution

- Definition: A vector is Gaussian if any linear combination of its components is a Gaussian variable

Conditional distribution

- Definition: A vector is Gaussian if any linear combination of its components is a Gaussian variable
- Property: When two components of a Gaussian vector are uncorrelated, they are independent

Conditional distribution

- Definition: A vector is Gaussian if any linear combination of its components is a Gaussian variable
- Property: When two components of a Gaussian vector are uncorrelated, they are independent
- Assumption: let consider a Gaussian vector

$$\mathbf{Y} = (Y(x_0), Y(x_1), \ldots, Y(x_n))$$

where $x_1, \ldots, x_n$ are the observation locations and x_0 the target location.

Conditional distribution

- Definition: A vector is Gaussian if any linear combination of its components is a Gaussian variable
- Property: When two components of a Gaussian vector are uncorrelated, they are independent
- Assumption: let consider a Gaussian vector

$$\mathbf{Y} = (Y(x_0), Y(x_1), \ldots, Y(x_n))$$

  where $x_1, \ldots, x_n$ are the observation locations and x_0 the target location.
- Conditional distribution: distribution of $Y(x_0)$ knowing

$$\text{Data} = \{Y(x_1) = y_1, \ldots, Y(x_n) = y_n\}?$$

Show that

$$Y(x_0) = Y^\star(x_0) + \sigma R$$

where

- $Y^\star(x_0)$ is the simple kriging of $Y(x_0)$ from the variables $Y(x_1), \ldots, Y(x_n)$
- σ^2 is the associated prediction variance ($Var(Y^\star(x_0) - Y(x_0))$)
- R is a Gaussian variable, centered and standardized, independent of any component $Y(x_i), i = 1, \ldots, n$

Conditional distributions

Gaussian case

Show that

$$Y(x_0) = Y^\star(x_0) + \sigma R$$

where

- $Y^\star(x_0)$ is the simple kriging of $Y(x_0)$ from the variables $Y(x_1), \ldots, Y(x_n)$
- σ^2 is the associated prediction variance ($Var(Y^\star(x_0) - Y(x_0))$)
- R is a Gaussian variable, centered and standardized, independent of any component $Y(x_i), i = 1, \ldots, n$

Proof: set

$$R = \frac{Y(x_0) - Y^\star(x_0)}{\sigma}$$

and check the hypothesis on R

Conditional distributions

Gaussian case

- Conclusion : the (conditional) distribution of $Y(x_0)$ knowing

$$\text{Data} = \{Y(x_1) = y_1, \ldots, Y(x_n) = y_n\}$$

  is Gaussian with mean $y^\star(x_0)$ and variance σ^2.

- One can deduce for any cutoff value y_c,

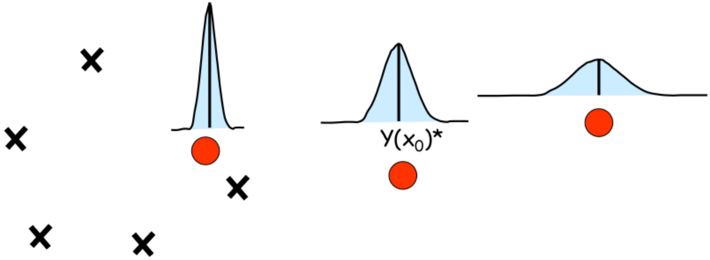
$$P(Y(x_0) > y_c | \text{Data}) = 1 - G\left(\frac{y_c - y^\star(x_0)}{\sigma}\right)$$

  where G is the standardized Gaussian c.d.f.

- the conditional variance σ^2 only depends on the locations

Conditional distributions

Gaussian case



Regression

Gaussian case

- Regression : find the function $r(Y(x_1), \ldots, Y(x_n))$ which minimizes the mean squared error:

$$E\left[(Y(x_0) - r(Y(x_1), \ldots, Y(x_n)))^2\right]$$

Regression

Gaussian case

- Regression : find the function $r(Y(x_1), \ldots, Y(x_n))$ which minimizes the mean squared error:

$$E\left[(Y(x_0) - r(Y(x_1), \ldots, Y(x_n)))^2\right]$$

- the solution is the conditional expectation

$$r(Y(x_1), \ldots, Y(x_n)) = E[Y(x_0) \,|\, Y(x_1), \ldots, Y(x_n)]$$

Regression

Gaussian case

- Regression : find the function $r(Y(x_1), \ldots, Y(x_n))$ which minimizes the mean squared error:

$$E\left[(Y(x_0) - r(Y(x_1), \ldots, Y(x_n)))^2\right]$$

- the solution is the conditional expectation

$$r(Y(x_1), \ldots, Y(x_n)) = E[Y(x_0)\,|\,Y(x_1), \ldots, Y(x_n)]$$

- It is an unbiased predictor (thanks to the law of the total expectation $E[E[Y|X]] = E[Y]$)

Regression

Gaussian case

- Regression : find the function $r(Y(x_1), \ldots, Y(x_n))$ which minimizes the mean squared error:

$$E\left[(Y(x_0) - r(Y(x_1), \ldots, Y(x_n)))^2\right]$$

- the solution is the conditional expectation

$$r(Y(x_1), \ldots, Y(x_n)) = E[Y(x_0) \,|\, Y(x_1), \ldots, Y(x_n)]$$

- It is an unbiased predictor (thanks to the law of the total expectation $E[E[Y|X]] = E[Y]$)
- Kriging is optimal in the Gaussian case

Regression

Gaussian case

- Regression : find the function $r(Y(x_1), \ldots, Y(x_n))$ which minimizes the mean squared error:

$$E\left[(Y(x_0) - r(Y(x_1), \ldots, Y(x_n)))^2\right]$$

- the solution is the conditional expectation

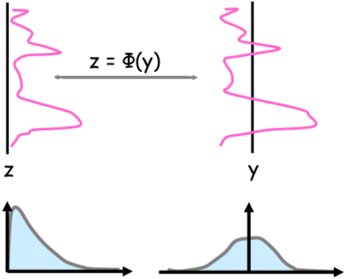
$$r(Y(x_1), \ldots, Y(x_n)) = E[Y(x_0)\,|\,Y(x_1), \ldots, Y(x_n)]$$

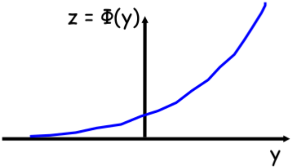
- It is an unbiased predictor (thanks to the law of the total expectation $E[E[Y|X]] = E[Y]$)
- Kriging is optimal in the Gaussian case
- It provides optimal predictor for any function f of Y:

$$E[f(Y(x_0)|\text{Data}] = \int_{-\infty}^{+\infty} f(y^\star(x_0) + \sigma t)g(t)dt$$

with g the standard Gaussian p.d.f ($g = G'$)

- The studied variable Z is often not Gaussian.
- Z will be modelled as a transformation by a function φ of a Gaussian variable Y
- φ is a bijective, increasing function (anamorphosis)





Example: log-normal case

We observe $Z(x_i) = z_i$ at locations $x_1, \ldots, x_n$ where

$$Z(x_i) = e^{\mu + \alpha Y(x_i)}$$

where Y is a standardized Gaussian function.

$$\text{Data} = \{Y(x_1) = y_1, \ldots, Y(x_n) = y_n\}$$

where

$$y_i = \frac{\log(z(x_i)) - \mu}{\alpha}$$

Example: log-normal case

We observe $Z(x_i) = z_i$ at locations $x_1, \ldots, x_n$ where

$$Z(x_i) = e^{\mu + \alpha Y(x_i)}$$

where Y is a standardized Gaussian function.

$$\text{Data} = \{Y(x_1) = y_1, \ldots, Y(x_n) = y_n\}$$

where

$$y_i = \frac{\log(z(x_i)) - \mu}{\alpha}$$

Then we have

$$P(Z(x_0) > z_c | \text{Data}) = 1 - G\left(\frac{y_c - y(x_0)^\star}{\sigma}\right)$$

with

$$y_c = \frac{\log(z_c) - \mu}{\alpha}$$

Conditional expectation

Log-normal case

- Prediction

$$Z_0^\star = E[Z(x_0)|\mathrm{Data}] = e^{\mu+\alpha Y^\star(x_0)+\frac{\alpha^2\sigma^2}{2}}$$

which is not the transform-back of the kriging!

Conditional expectation

Log-normal case

- Prediction

$$Z_0^\star = E[Z(x_0)|\mathrm{Data}] = e^{\mu+\alpha Y^\star(x_0)+\frac{\alpha^2\sigma^2}{2}}$$

  which is not the transform-back of the kriging!

- Error

$$E((Z(x_0)-Z_0^\star)^2|\mathrm{Data}) = \mathrm{Var}(Z(x_0)|\mathrm{Data}) = (Z_0^\star)^2(e^{\alpha^2\sigma^2}-1)$$

  which depends on the data values.

- Prediction

$$Z_0^\star = E[Z(x_0)|\mathrm{Data}] = e^{\mu+\alpha Y^\star(x_0)+\frac{\alpha^2\sigma^2}{2}}$$

  which is not the transform-back of the kriging!

- Error

$$E((Z(x_0)-Z_0^\star)^2|\mathrm{Data}) = \mathrm{Var}(Z(x_0)|\mathrm{Data}) = (Z_0^\star)^2(e^{\alpha^2\sigma^2}-1)$$

  which depends on the data values.

Proof:

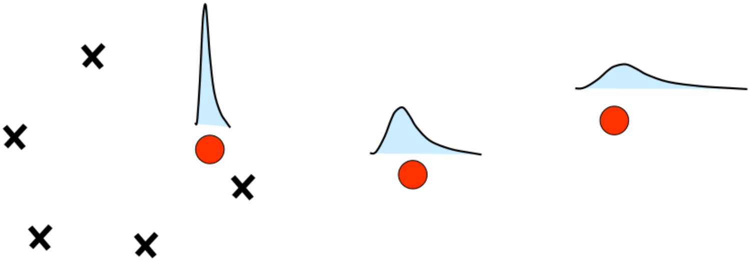
Use

$$E[e^{\lambda Y}] = e^{\frac{\lambda^2}{2}}$$

for a Gaussian variable Y with mean 0 and variance 1.

Conditional distributions

Log-normal case



Gaussian anamorphosis

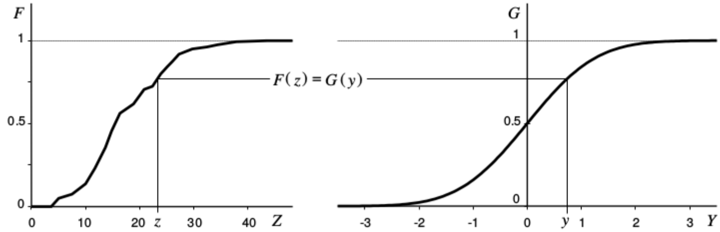
Transformation of the data values

When suitable transformation is not found :

- Compute the gaussian scores from the data z_i

$$y_i = G^{-1}(F(z_i))$$

  where F is the empirical cumulative distribution function of Z
- Fit a function φ (*anamorphosis*) $Z = \varphi(Y)$



Exercise

- Z is a random variable with c.d.f F.
- Compute the cdf H of $Y = G^{-1}(F(Z))$

Exercise

- Z is a random variable with c.d.f F.
- Compute the cdf H of $Y = G^{-1}(F(Z))$
- Solution

$$
\begin{aligned}
H(y) &= P(Y \leq y) \\
&= P(G^{-1}(F(Z)) \leq y) \\
&= P(F(Z) \leq G(y)) \\
&= P(Z \leq F^{-1}(G(y))) \\
&= F(F^{-1}(G(y)) \\
&= G(y)
\end{aligned}
$$

Summary

- Compute the Gaussian scores $y_1, \ldots, y_n$
- Fit an anamorphosis function φ
- Compute simple kriging $y^\star(x_0)$ of the gaussian scores and the associated estimation variance σ^2
- Compute

$$P(Z(x_0) \geq z_c|\text{Data}) = 1 - G\left(\frac{\varphi^{(-1)}(z_c) - y^\star(x_0)}{\sigma}\right)$$

- Estimation by conditional expectation

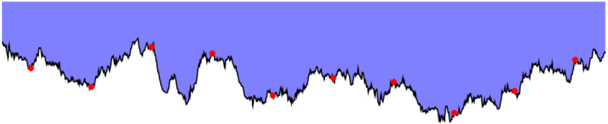
$$Z^\star(x_0) = \int \varphi(y^\star(x_0) + \sigma t)g(t)dt$$

- Error

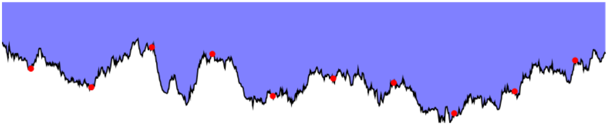
$$\text{Var}(Z(x_0)|\text{Data})$$

Why simulations?

Example: a submarine cable has to be set on the seabed between Lisbonne and New-York.
How to predict its length from the bathymetry sampled every 100m.
Uncertainty?



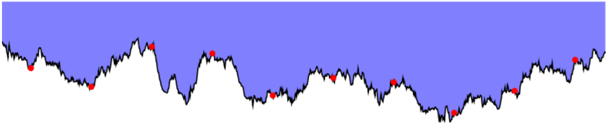
Natural idea



Measuring the length of the predicted (kriged) bathymetry

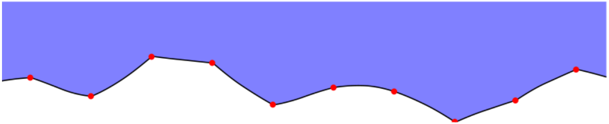
$$\hat{l} = \int_a^b \sqrt{1 + [\hat{z}(x)']^2} dx$$

Natural idea



Measuring the length of the predicted (kriged) bathymetry

$$\hat{l} = \int_a^b \sqrt{1 + [\hat{z}(x)']^2} dx$$

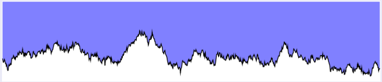


Systematic under-estimation: predicted trajectory is much smoother than the actual one

Idea: Monte-Carlo simulations

Idea: Monte-Carlo simulations

Unconditional simulations

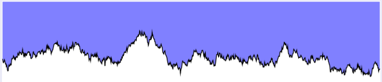


Output: A set $t_1, \ldots, t_N$ of cable lengths drawn from the model

$$P(T(Z) \leq t)$$

Idea: Monte-Carlo simulations

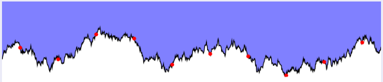
Unconditional simulations



Output: A set $t_1, \ldots, t_N$ of cable lengths drawn from the model

$$P(T(Z) \leq t)$$

Conditional simulations



Output: A set $t_1, \ldots, t_N$ of cable lengths drawn from the conditional model

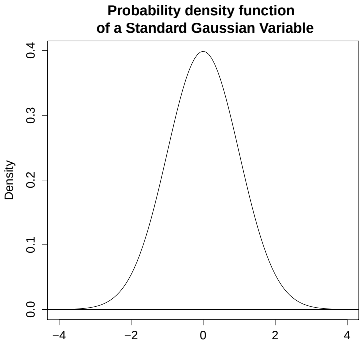
$$P(T(Z) \leq t \mid Z(x_1) = z_1, \ldots, Z(x_n) = z_n)$$

Gaussian random variable

A random variable Y is Gaussian with mean m and variance σ^2 if it has the following probability density function (p.d.f)

$$g_{m,\sigma^2}(y) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(y-m)^2}{2\sigma^2}\right), \qquad y \in \mathbb{R}$$

$$Y \sim \mathcal{N}(m, \sigma^2)$$



Simulation of a Gaussian variable

Box-Muller algorithm

- Let two independent random variables

$$U_1 \sim \mathcal{U}[0,1] \text{ and } U_2 \sim \mathcal{U}[0,1]$$

Simulation of a Gaussian variable

Box-Muller algorithm

- Let two independent random variables

$$U_1 \sim \mathcal{U}[0,1] \text{ and } U_2 \sim \mathcal{U}[0,1]$$

- then

$$Y_1 = \sqrt{-2\log U_1}\cos(2\pi U_2)$$

and

$$Y_2 = \sqrt{-2\log U_2}\cos(2\pi U_1)$$

are independent Gaussian random variables $\mathcal{N}(0,1)$

Simulation of a Gaussian variable

Box-Muller algorithm

- Let two independent random variables

$$U_1 \sim \mathcal{U}[0,1] \text{ and } U_2 \sim \mathcal{U}[0,1]$$

- then

$$Y_1 = \sqrt{-2\log U_1}\cos(2\pi U_2)$$

  and

$$Y_2 = \sqrt{-2\log U_2}\cos(2\pi U_1)$$

  are independent Gaussian random variables $\mathcal{N}(0,1)$

- Remark

$$b + aY_i \sim \mathcal{N}(b, a^2).$$

Central-limit theorem

- $(X_n)_{n\in\mathbb{N}^\star}$ is a sequence of independent and identically distributed $\mathbb{R}$-valued random variables (with mean m and finite variance σ^2).
- Empirical mean

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

- Then, the random variable

$$\sqrt{n}\frac{(\bar{X}_n - m)}{\sigma}$$

converges in distribution to a normal $\mathcal{N}(0,1)$.

Properties

1) Linear combination of the components

- $\mathbf{X}$ is a n-gaussian vector with mean $\mathbf{m}$ and covariance matrix M
- A is a matrix of size (p, n)
- $\mathbf{b}$ is a vector of size p

Properties

1) Linear combination of the components

- $\mathbf{X}$ is a n-gaussian vector with mean $\mathbf{m}$ and covariance matrix M
- A is a matrix of size (p, n)
- $\mathbf{b}$ is a vector of size p

Then $A\mathbf{X} + \mathbf{b}$ is a gaussian vector with

Properties

1) Linear combination of the components

- $\mathbf{X}$ is a n-gaussian vector with mean $\mathbf{m}$ and covariance matrix M
- A is a matrix of size (p, n)
- $\mathbf{b}$ is a vector of size p

Then $A\mathbf{X} + \mathbf{b}$ is a gaussian vector with

- $E[A\mathbf{X} + \mathbf{b}] = A\mathbf{m} + \mathbf{b}$
- $\mathrm{Var}(A\mathbf{X} + \mathbf{b}) = AMA^t$

Application: simulation of a Gaussian vector

Aim:

- Simulate a Gaussian vector $\mathbf{Y} = (Y_1, \ldots, Y_n)$ with mean $\mathbf{m}$ and a covariance matrix Σ.

Application: simulation of a Gaussian vector

Aim:

- Simulate a Gaussian vector $\mathbf{Y} = (Y_1, \ldots, Y_n)$ with mean $\mathbf{m}$ and a covariance matrix Σ.

Ingredient:

- If Σ is positive-definite, there is a matrix A of size $n \times n$ such as $\Sigma = AA^t$ ("square-root of Σ")
- Example: Cholesky decomposition (LL^t), ...

Application: simulation of a Gaussian vector

Aim:

- Simulate a Gaussian vector $\mathbf{Y} = (Y_1, \ldots, Y_n)$ with mean $\mathbf{m}$ and a covariance matrix Σ.

Ingredient:

- If Σ is positive-definite, there is a matrix A of size $n \times n$ such as $\Sigma = AA^t$ ("square-root of Σ")
- Example: Cholesky decomposition (LL^t), ...

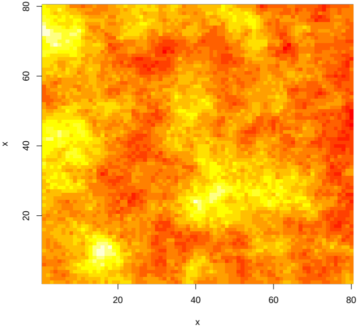
Algorithm:

1. Compute A, a "square-root" of Σ (i.e $\Sigma = AA^t$)
2. Simulate n independent Gaussian normal variables $\mathbf{X} = (X_1, \ldots, X_n)$.
3. Return

$$\mathbf{Y} = \mathbf{m} + A\mathbf{X}$$

2D EXAMPLE

A Gaussian random function with $C(h) = e^{-h/10}$ on a 80× 80 regular grid.



Properties

2) Conditional distribution

Let consider a Gaussian vector

$$Z = (Y, X) = (Y_1, \dots, Y_n, X_1, \dots, X_p)$$

with

$$E[Z] = \begin{pmatrix} m_Y \\ m_X \end{pmatrix}$$

and

$$\mathrm{Cov}(Z) = \begin{pmatrix} \Sigma_{YY} & \Sigma_{XY}^t \\ \Sigma_{XY} & \Sigma_{XX} \end{pmatrix} = \begin{pmatrix} \mathrm{Cov}(Y, Y) & \mathrm{Cov}(Y, X) \\ \mathrm{Cov}(X, Y) & \mathrm{Cov}(X, X) \end{pmatrix}$$

Conditional distribution of Y knowing $X = \mathbf{x}$?

Properties

2) Conditional distribution

$$Y = Y^\star + Y - Y^\star$$

with

$$Y^\star = m_Y + \Sigma_{XY}^t \Sigma_{XX}^{-1}(X - m_X)$$

- $Y^\star$ contains the kriging of the components of Y
- The vector $(X, Y - Y^\star)$ is Gaussian
- $\mathrm{Cov}(X, Y - Y^\star) = 0$ (independence)
- $E[Y - Y^\star] = 0$
- $\mathrm{Var}(Y - Y^\star) = \Sigma_{YY} - \Sigma_{XY}^t \Sigma_{XX}^{-1} \Sigma_{XY}$

Relationship between X and Y

$$Y = f(X) + R$$

- $R \sim \mathcal{N}$
- $R \perp\!\!\!\perp X$

The distribution of Y conditionally to $X = \mathbf{x}$ is Gaussian

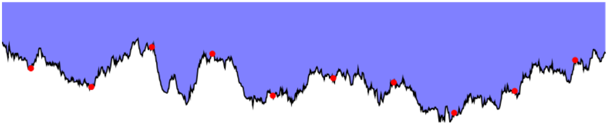
- with mean
$$m_Y + \Sigma_{XY}^t \Sigma_{XX}^{-1}(\mathbf{x} - m_X) \tag{1}$$
- and variance
$$\Sigma_{YY} - \Sigma_{XY}^t \Sigma_{XX}^{-1} \Sigma_{XY} \tag{2}$$

Remark: the conditional variance of Y does not depend on X.

Example

The problem

- A submarine cable has to be set on the seabed between two coasts separated of 1km. How to predict its length from the bathymetry sampled every 50m. Uncertainty?



Example

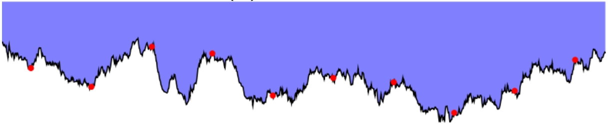
The problem

- The model for the bathymetry is a Gaussian random function with mean

$$m = -400 \text{ meters}$$

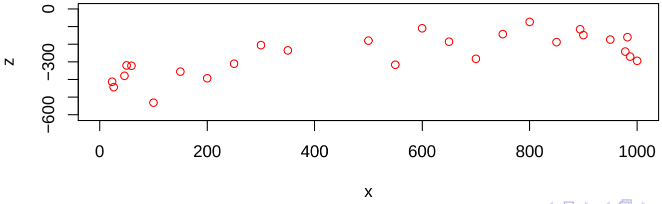
and with a covariance function

$$C(h) = 16000 e^{-\frac{h}{50 \text{ meters}}}$$



Example
Data

- The data set

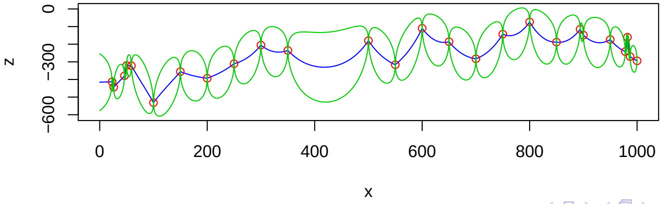


Example

Pointwise conditional distribution

- Conditional distribution of the bathymetry $Y(x_0)$ knowing $Y(x_1), \dots Y(x_n)$:

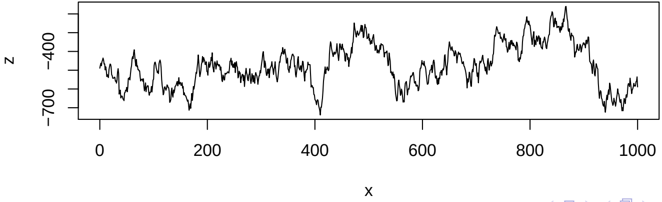
$$\mathcal{N}(Y^K(x_0), \sigma_K)$$



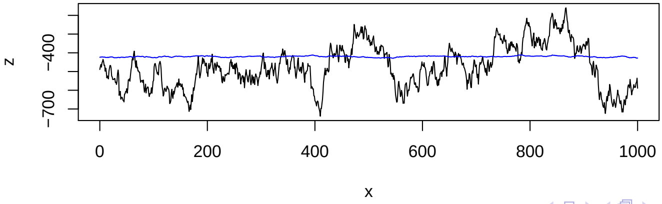
Example

Non conditional simulations

- One non conditional simulation



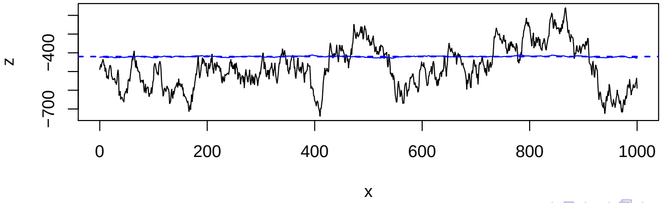
- One non conditional simulation
- Average over 1000 simulations



Example

Non conditional simulations

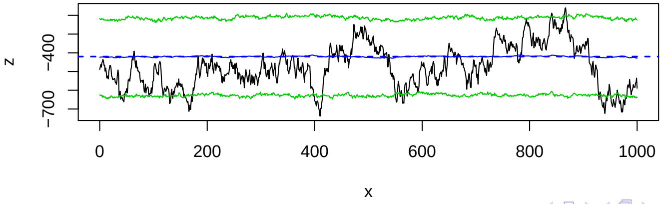
- One non conditional simulation
- Comparison with the theoretical mean



Example

Non conditional simulations

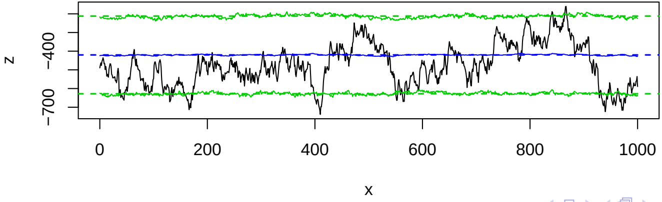
- One non conditional simulation
- Empirical pointwise quantiles of 1000 simulations



Example

Non conditional simulations

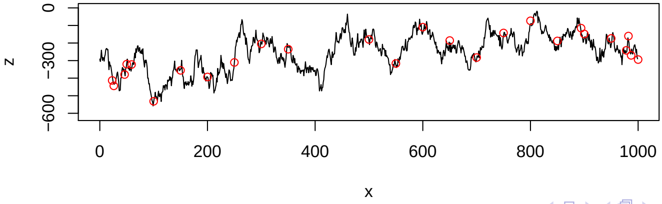
- One non conditional simulation
- Comparison with the theoretical quantiles



Example

Conditional simulations

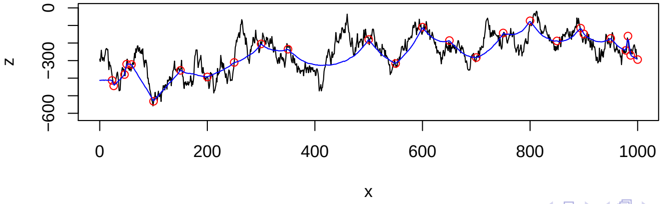
- One conditional simulation



Example

Conditional simulations

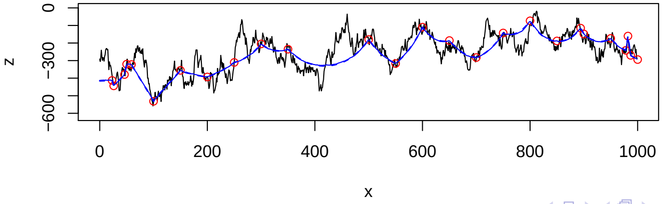
- One conditional simulation
- Average over 1000 simulations



Example

Conditional simulations

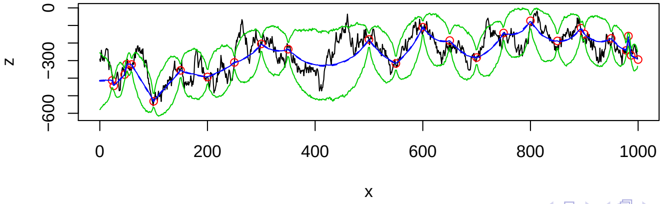
- One conditional simulation
- Comparison with the theoretical mean



Example

Conditional simulations

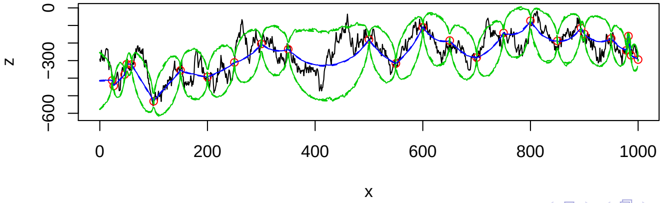
- One conditional simulation
- Empirical pointwise quantiles of 1000 simulations



Example

Conditional simulations

- One conditional simulation
- Comparison with the theoretical quantiles

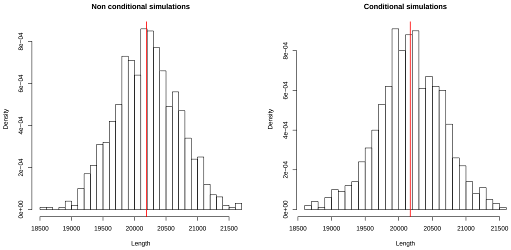


Conclusion

- Actual length: 20426 m
- Length of the kriging: 2708.6 m
- Average length of the non-conditional simulations
  $E_{\text{model}}[l] = 20192.6\text{m}$
- Average length of the conditional simulations
  $E_{\text{model}}[l|data] = 20170.8\text{m}$

Conclusion

- Actual length: 20426 m
- Length of the kriging: 2708.6 m
- Average length of the non-conditional simulations $E_{\text{model}}[l] = 20192.6\text{m}$
- Average length of the conditional simulations $E_{\text{model}}[l|data] = 20170.8\text{m}$



- Standard dev. of the non conditional simulations $\sqrt{\text{V}_{\text{model}}[l]} = 497.6\text{m}$
- Standard dev. of the conditional simulations $\sqrt{\text{V}_{\text{model}}[l|data]} = 486.5\text{m}$

Simulation of a Gaussian random function with a given covariance function

- Decomposition of the covariance matrix ($n < 10000$)
- Otherwise, functional extension of the central-limit theorem (spectral method, turning bands,...)

Functional extension of the central-limit theorem

Hypothesis

$(Z_n, n \in \mathbb{N}^\star)$ is a sequence of second-order stationary random functions with mean m, variance σ^2 and covariance function $C = \sigma^2\rho$, and if

$$\bar{Z}_n = \frac{Z_1 + \cdots + Z_n}{n}$$

Conclusion

When $n \to \infty$, the random function

$$Y^{(n)} = \sqrt{n}\frac{\bar{Z}_n - m}{\sigma}$$

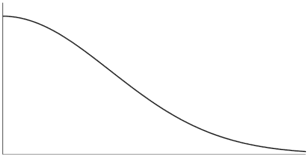
converges to a Gaussian random function with mean 0, variance 1 and covariance function ρ.

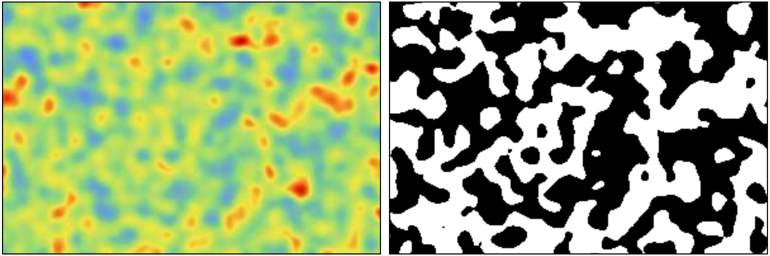
Use in simulation algorithms

- Aim: simulate second-order stationary Gaussian random functions with mean m and covariance C ($m = 0$ and $\sigma^2 = 1$)
- Simulate simple processes with covariance C
- Average these processes to obtain a Gaussian random function with a covariance C
- 1 - Spectral method
- 2 - Turning bands
- 3 - *Ad hoc* methods

Gaussian covariance

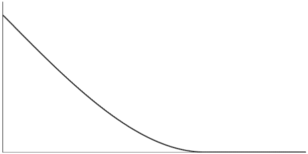
$$C(h) = \exp\left(-\frac{|h|^2}{a^2}\right)$$

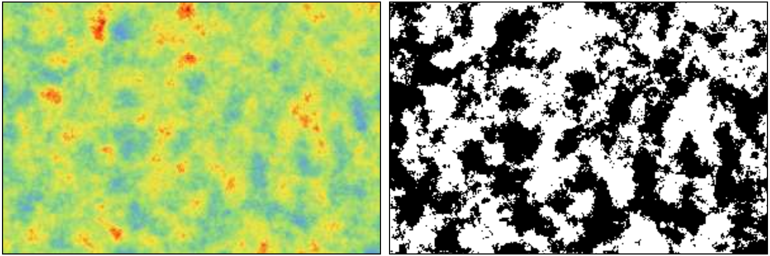




Spherical covariance

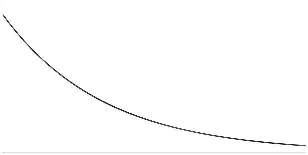
$$C(h) = \left(1 - \frac{3}{2}\frac{|h|}{a} + \frac{1}{2}\frac{|h|^3}{a^3}\right) 1_{|h| \leq a}$$

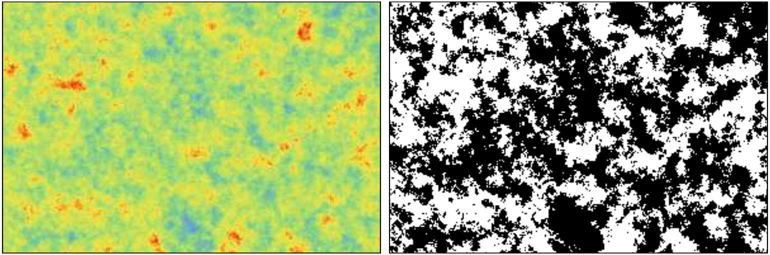




Exponential covariance

$$C(h) = \exp\left(-\frac{|h|}{a}\right)$$





Bochner theorem

If C is continuous at 0, then C is a covariance function if and only if there exists a probability measure F (the spectral measure) such as

$$C(h) = \int_{\mathbb{R}^d} e^{i\omega^t.h} dF(\omega)$$

When $C \in L^2(\mathbb{R}^d)$

$$C(h) = \int_{\mathbb{R}^d} e^{i\omega^t.h} f(\omega) d\omega$$

with $f = F'$ is the spectral density C

$$f(w) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} e^{-i\omega^t.h} C(h) dh$$

Consequence

Since a covariance function is an even function we have

$$C(h) = \int_{\mathbb{R}^d} \cos(\omega^t.h) f(\omega) d\omega$$

Consequence

Since a covariance function is an even function we have

$$C(h) = \int_{\mathbb{R}^d} \cos(\omega^t . h) f(\omega) d\omega$$

Property

If

- $\Omega \sim F$ (with density f)
- $\Phi \sim \mathcal{U}([0, 2\pi))$
- Ω and Φ are independent

Then, the random function defined for all $x \in \mathbb{R}^d$ by

$$Z(x) = \sqrt{2} \cos\left(\Omega^t . x + \Phi\right)$$

is second-order stationary with mean 0 and covariance function C.

Spectral method

Algorithm:

(i) Simulate n independent random vectors of $\mathbb{R}^d$

$$\omega_1, \ldots, \omega_n \sim F$$

and n independent phases

$$\phi_1, \ldots, \phi_n \sim \mathcal{U}([0, 2\pi))$$

(ii) For each point $x \in \mathbb{R}^d$*, compute*

$$y(x) = \frac{\sqrt{2}}{\sqrt{n}} \sum_{j=1}^{n} \cos(\omega_j^t . x + \phi_j)$$

Example

Gaussian covariance

$$C(h) = \exp\left(-\frac{|h|^2}{a^2}\right)$$

Spectral measure of C:
Since C is L^2, the spectral measure F of C is given by:

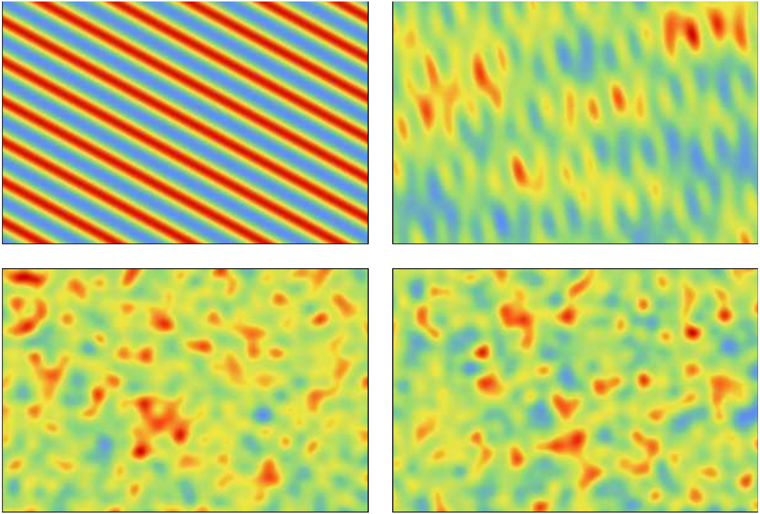
Spectral density

$$f(\omega) = \left(\frac{a}{2\sqrt{\pi}}\right)^d \exp\left(-\frac{a^2|\omega|^2}{4}\right)$$

To simulate $\omega = (u_1, \ldots, u_d)$ according to F, one have to simulate d independent realizations

$$u_1, \ldots, u_d \sim \mathcal{N}(0, 2/a^2)$$

Illustration



Simulation with $1, 10, 100$ and 1000 basic functions

Remarks

- When the covariance function is not smooth close to 0, the spectral density can be heavy-tailed
- Therefore, a large number of basic simulations is required in order to have a representative sample from the spectral measure: lack of efficiency
- Solution: add stochasticity by using turning bands method

Turning bands

Bochner theorem:

$$C(h) = \int_{\mathbb{R}^d} e^{i\omega^t . h} dF(\omega)$$

In spherical coordinates:

Let $\omega = r\vec{\theta}$ with $r \in \mathbb{R}$ and $\vec{\theta} \in S^+_{d-1}$ (half unit sphere of $\mathbb{R}^d$). Then

$$C(h) = \int_{S^+_{d-1}} \int_{\mathbb{R}} e^{ir\vec{\theta}^t . h} dF_{\vec{\theta}}(r) d\varpi(\vec{\theta}) = \int_{S^+_{d-1}} C_{\vec{\theta}}(\vec{\theta}^t . h) d\varpi(\vec{\theta})$$

where

$$C_{\vec{\theta}}(u) = \int_{\mathbb{R}} e^{iru} dF_{\vec{\theta}}(r)$$

are some unidimensional covariances.

Turning band algorithm

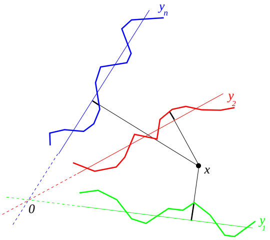
(i) Simulate n independent directions of $\mathbb{R}^d$,

$$\vec{\theta}_1, \ldots, \vec{\theta}_n \sim \varpi$$

(ii) simulate $z_1, \ldots, z_n$ *some realizations of n independent stochastic processes* $Z_1, \ldots, Z_n$ *all with mean* 0 *and respective covariance functions* $C_{\vec{\theta}_1}, \ldots, C_{\vec{\theta}_n}$

(iii) For each point $x \in \mathbb{R}^d$, *compute*

$$y(x) = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} z_j(x^t.\vec{\theta}_j)$$



In the 3D isotropic case

- the covariance function C can be written $C(h) = C_3(|h|)$ (for $h \in \mathbb{R}^3$)

In the 3D isotropic case

- the covariance function C can be written $C(h) = C_3(|h|)$ (for $h \in \mathbb{R}^3$)
- Since the Fourier transform of a radial function is radial,

In the 3D isotropic case

- the covariance function C can be written $C(h) = C_3(|h|)$ (for $h \in \mathbb{R}^3$)
- Since the Fourier transform of a radial function is radial,
  - ϖ is the uniform measure on $\mathcal{S}_2^+$

In the 3D isotropic case

- the covariance function C can be written $C(h) = C_3(|h|)$ (for $h \in \mathbb{R}^3$)
- Since the Fourier transform of a radial function is radial,
  - ϖ is the uniform measure on $\mathcal{S}_2^+$
  - all the covariances $C_{\vec{\theta}}$ are equal to the same unidimensional covariance function C_1.

In the 3D isotropic case

- the covariance function C can be written $C(h) = C_3(|h|)$ (for $h \in \mathbb{R}^3$)
- Since the Fourier transform of a radial function is radial,
  - ϖ is the uniform measure on $\mathcal{S}_2^+$
  - all the covariances $C_{\vec{\theta}}$ are equal to the same unidimensional covariance function C_1.
- C_1 and C_3 are linked by:

$$C_1(r) = \frac{d}{dr}\left(rC_3(r)\right) \qquad r > 0$$

Elements of proof

We have

$$
\begin{aligned}
C_3(|h|) &= C(h) = E[Z(x+h)Z(x)] \\
&= E[E[Z(x+h)Z(x)|\vec{\theta}]] = E[C_1(h^t.\vec{\theta})] = \int_{S_2^+} C_1(\vec{\theta}^t.h)d\varpi(\vec{\theta})
\end{aligned}
$$

where ϖ is the uniform measure on S_2^+.

Elements of proof

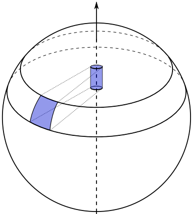
We have

$$C_3(|h|) = C(h) = E[Z(x+h)Z(x)]$$
$$= E[E[Z(x+h)Z(x)|\vec{\theta}]] = E[C_1(h^t.\vec{\theta})] = \int_{S_2^+} C_1(\vec{\theta}^t.h)d\varpi(\vec{\theta})$$

where ϖ is the uniform measure on S_2^+.
It follows

$$C_3(r) = \int_0^1 C_1(zr)dz = \frac{1}{r}\int_0^r C_1(u)du$$

as the projection on a radius of a uniform point on $\mathcal{S}_2^+$ is uniform.



Algorithm

(i) Simulate n independent uniform directions $\vec{\theta}_1, \ldots, \vec{\theta}_n$ of $\mathbb{R}^d$
(ii) simulate $z_1, \ldots, z_n$ some realizations of n independent stochastic processes $Z_1, \ldots, Z_n$ all with mean 0 and covariance function C_1
(iii) For each point $x \in \mathbb{R}^d$, compute

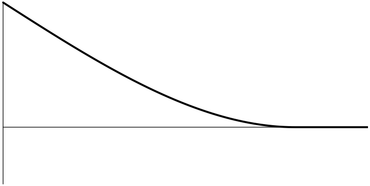
$$\frac{1}{\sqrt{n}} \sum_{j=1}^{n} z_j(x^t.\vec{\theta}_j)$$

Spherical covariance (1/3)

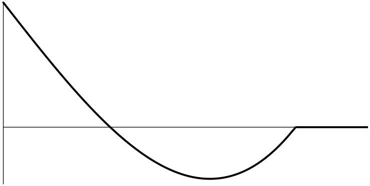
From 3D to 1D:

$$C(r) = \left(1 - \frac{3}{2}\frac{r}{a} + \frac{1}{2}\frac{r^3}{a^3}\right) 1_{r \leq a} \Longrightarrow C_1(r) = \left(1 - 3\frac{r}{a} + 2\frac{r^3}{a^3}\right) 1_{r \leq a}$$

Shape of C_3:



Shape of C_1:



Spherical covariance (2/3)

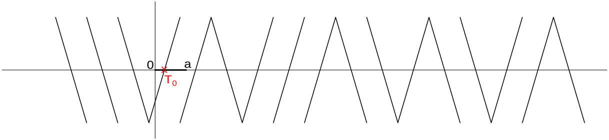
Algorithm to generate a stochastic process with covariance C_1:

(i) Simulate T_0 *uniform on* $[0, a]$ *and set* $T_k = T_0 + ka$ *for* $k \in \mathbb{Z}$
(ii) Split the line in intervals $I_k = [T_k - \frac{a}{2}, T_k + \frac{a}{2})$, $k \in \mathbb{Z}$
(iii) For all $t \in \mathbb{R}$, *compute*

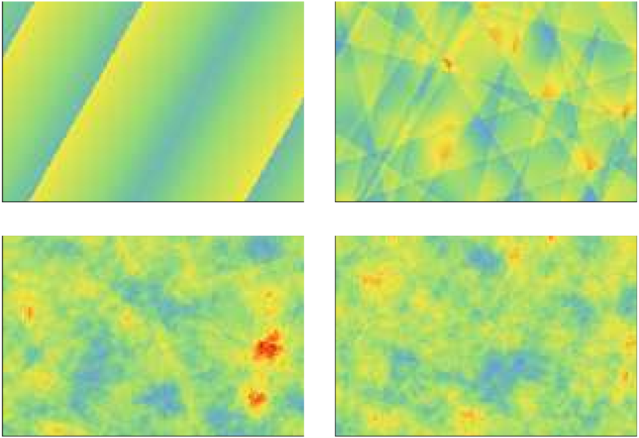
$$z(t) = \sum_{k \in \mathbb{Z}} \varepsilon_k 2 \frac{\sqrt{3}}{a} (t - T_k) 1_{I_k}(t)$$

where the ε_k *are independent random signs (independent of* T_0*):*
$P(\varepsilon_k = 1) = P(\varepsilon_k = -1) = 1/2$

Example:



Spherical covariance (3/3)



Simulation with $1, 10, 100$ and 1000 bands

Ad hoc METHOD: EXPONENTIAL COVARIANCE

Remark:

$$\exp\left(-\frac{|h|}{a}\right) = \int_0^{\infty} C_{\mathrm{spher}}\left(\frac{|h|}{au}\right)\omega(u)du$$

- C_{spher} is the spherical covariance with scale parameter 1
- $\omega(u) = \frac{1}{3}e^{-u}u(1+u)$ is a p.d.f (mixture of Gamma distributions)

Ad hoc method: exponential covariance

Remark:

$$\exp\left(-\frac{|h|}{a}\right) = \int_0^\infty C_{\text{spher}}\left(\frac{|h|}{au}\right)\omega(u)du$$

- C_{spher} is the spherical covariance with scale parameter 1
- $\omega(u) = \frac{1}{3}e^{-u}u(1+u)$ is a p.d.f (mixture of Gamma distributions)

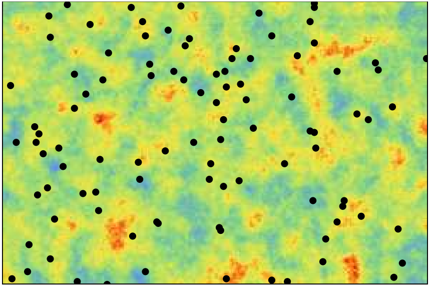
Algorithm:

- *Simulate n independent and uniform directions* $\theta_1, \ldots, \theta_n$ *and* $u_1, \ldots, u_n \sim \omega$
- *simulate* $z_1, \ldots, z_n$ *some realizations of n independent stochastic processes* $Z_1, \ldots, Z_n$*, each with mean* 0 *and covariance function* C_1 *(but with scale parameters* $au_i, i = 1, \ldots, n$*)*
- *For all* $x \in \mathbb{R}^d$*, compute*

$$y(x) = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} z_j(x^t.\theta_j)$$

Conditional simulations

How to simulate a realization $\{y(x), x \in \mathbb{R}^d\}$ of a second order Gaussian random function $\{Y(x), x \in \mathbb{R}^d\}$ with mean m and covariance function C with respect to the data $\{Y(x_i) = y_i, i = 1, \ldots, n\}$?



Principle

Let

$$Y(x) = Y^{SK}(x) + Y(x) - Y^{SK}(x)$$

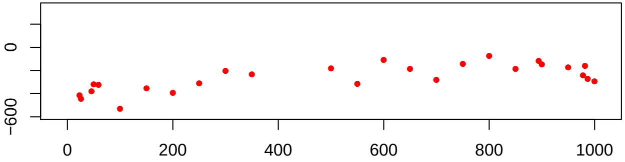
where

$Y^{SK}(x) = m + \sum_{j=1}^{n} \lambda_j(x)[Y(x_j) - m]$ simple kriging

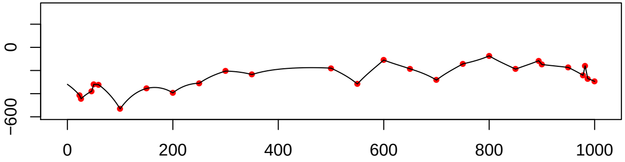
$Y(x) - Y^{SK}(x)$ kriging residuals

Y^{SK} and $Y - Y^{SK}$ are two independent Gaussian random functions

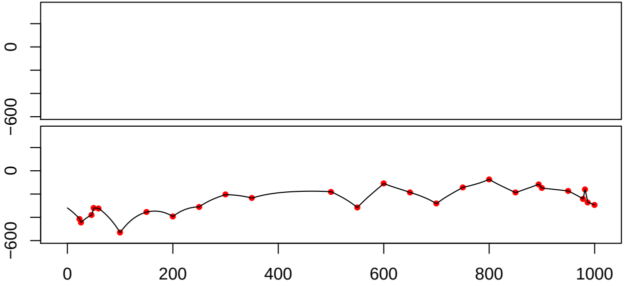
Principle



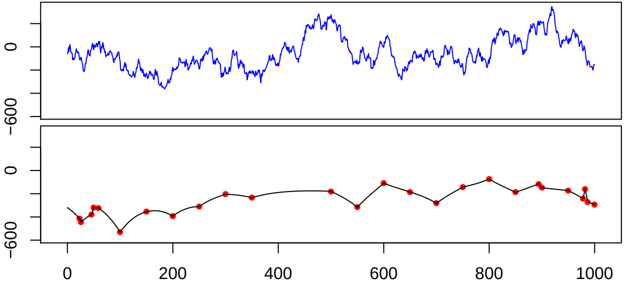
Principle



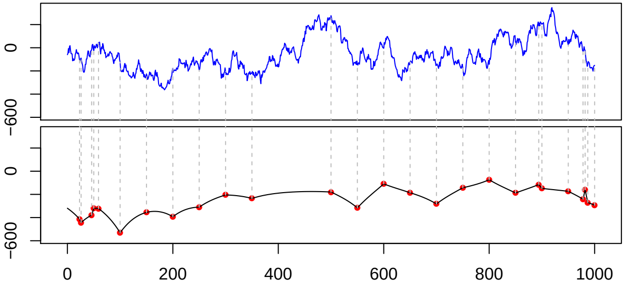
Principle



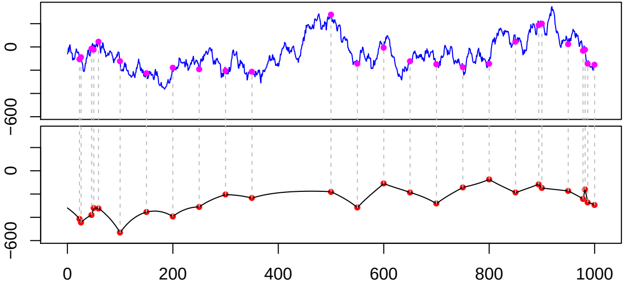
Principle



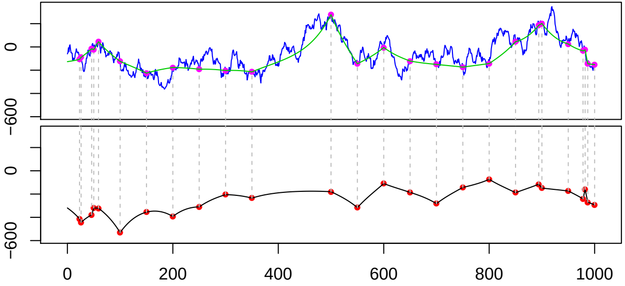
Principle



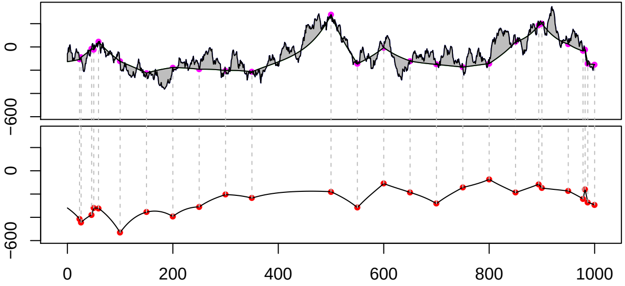
Principle



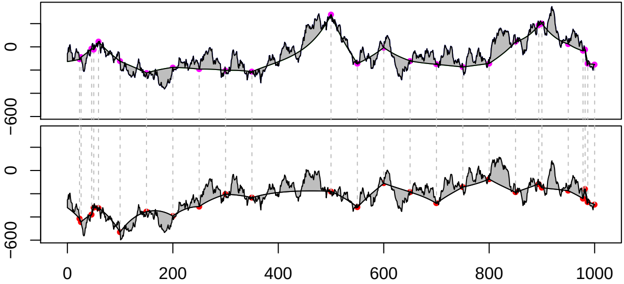
Principle



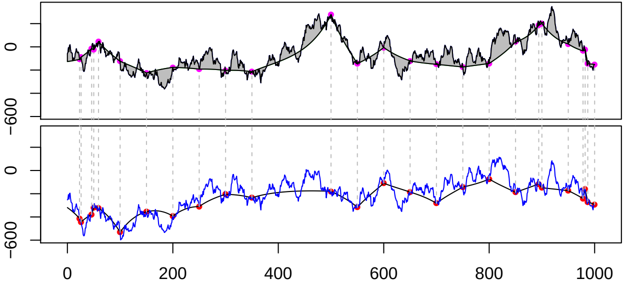
Principle



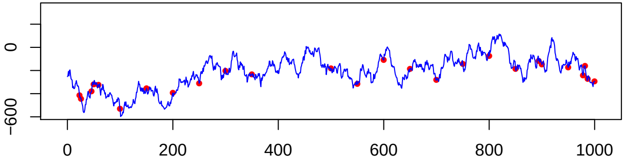
Principle



Principle



Principle



Algorithm

(i) make an unconditional simulation $\{s(x), x \in \mathbb{R}^d\}$ *and set* $s_j = s(x_j)$.

(ii) for all $x \in \mathbb{R}^d$, *compute the kriging weights* $(\lambda_j(x), j = 1, \ldots, n)$

(iii) set

$$\begin{aligned} y^{CS}(x) &= y^{SK}(x) + s(x) - s^{SK}(x) \\ &= s(x) + \sum_{j=1}^{n} \lambda_j(x)(y_j - s_j) \end{aligned}$$

Verification

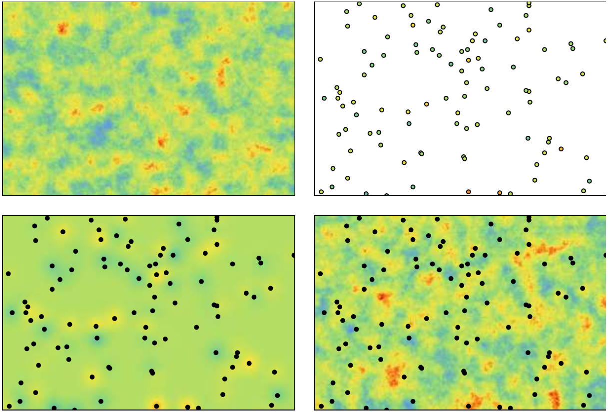
- If $x = x_i$, then $y^{CS}(x_i) = y_i + s(x_i) - s(x_i) = y_i$

  The conditioning data are honored

- If $C(x - x_\alpha) \simeq 0$ for all $i = 1, \ldots, n$, then
  $y^{CS}(x) \simeq m + s(x) - m = s(x)$

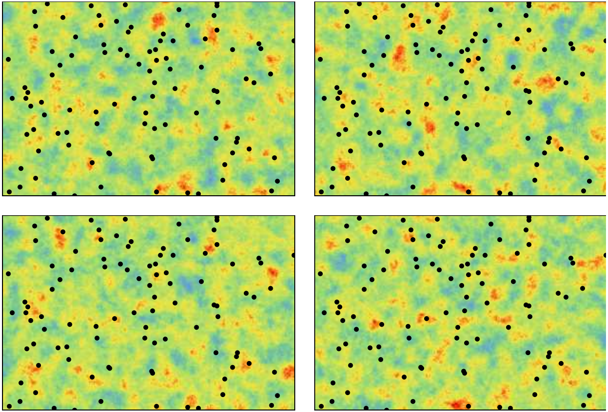
  For points distant from the data, the simulation is unconditional

Illustration



Simulation (TL), Synthetic data (TR), Simple kriging (BL), a conditional simulation (BR)

4 other conditional simulations



Gaussian anamorphosis

- Limitation of the Gaussian model:
  - Symetrical distribution
  - Inadapted to model data with heavy tail histogram
  - Conditional variance does not depend on the data (homoscedasticity)
- Anamorphosis: bijective function φ defined on $\mathbb{R}$
- Model: data are a realization of a second order stationary Gaussian random function transformed by φ
- Example : The log-normal model

$$\varphi(Y) = \exp(\mu + \sigma Y)$$

Conditional simulation of a transformed Gaussian random function

Model

$(z(x_1), \ldots, z(x_n))$ is a realization of $(Z(x), x \in \mathbb{R}^d)$, a transformed Gaussian random function:

- $Z(x) = \varphi(Y(x))$ for all $x \in \mathbb{R}^d$
- $(Y(x), x \in \mathbb{R}^d)$ a second order stationary Gaussian random function with correlation function C

How to simulate a realization $(z(x), x \in \mathbb{R}^d)$ from $(Z(x), x \in \mathbb{R}^d)$ such as $Z(x_i) = z(x_i)$ for $i = 1, \ldots, n$?

Algorithm

(i) transform the data in the Gaussian scale

$$y_i = \varphi^{-1}(z(x_i))$$

(ii) Perform a conditional simulation $\{y(x), x \in \mathbb{R}^d\}$ from the Gaussian random function $(Y(x), x \in \mathbb{R}^d)$ which honors the data in the Gaussian case

$$Y(x_i) = y_i$$

(iii) transform back the simulation for all $x \in \mathbb{R}^d$

$$z(x) = \varphi(y(x))$$