Prediction by conditional simulation

C. Lantuéjoul

MinesParisTech

christian.lantuejoul@mines-paristech.fr

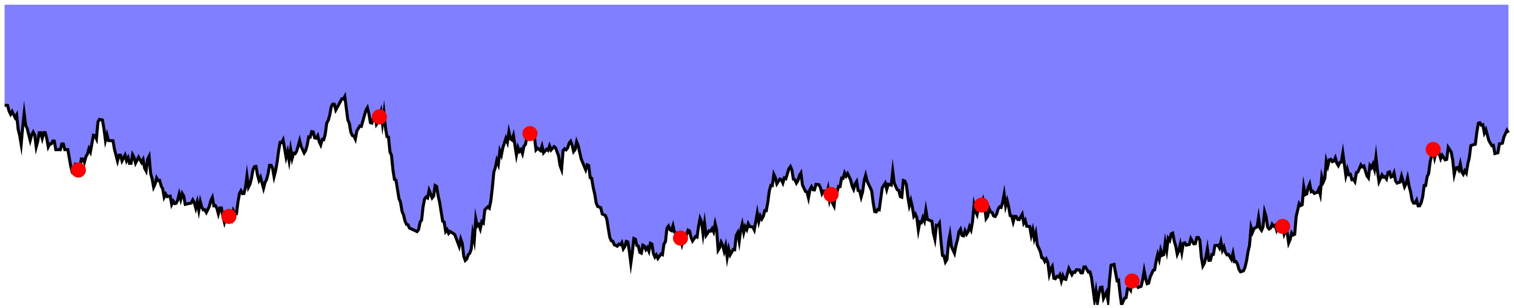
Introductive example

Problem:

A submarine cable has to be laid on the see floor between points a and b. We want to predict its length

$$\ell = \int_a^b \sqrt{1 + [y'(x)]^2}\, dx$$

starting from the sea-bed topography $\big(y(x), a \leq x \leq b\big)$ along a trajectory that has been sampled every 100m.



Submarine trajectory and samples

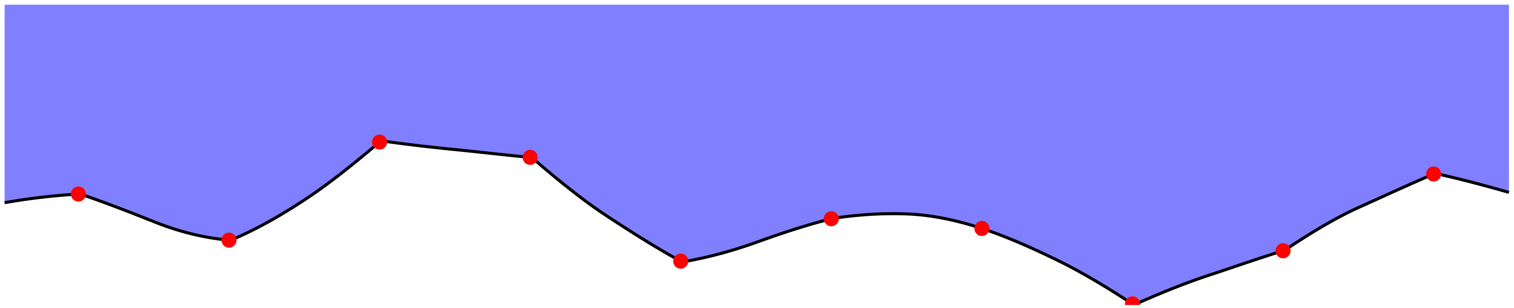
Introductive example (2)

A natural idea:

Predict the cable length as the length of the predicted trajectory $\big(\hat{y}(x), a \leq x \leq b\big)$

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

Unfortunately, the predicted value $\hat{\ell}$ is much smaller than the actual one ℓ.



Predicted submarine trajectory and samples

Introductive example (3)

Comments:

The prediction algorithm proposed is inefficient because of

– the data: no local information available

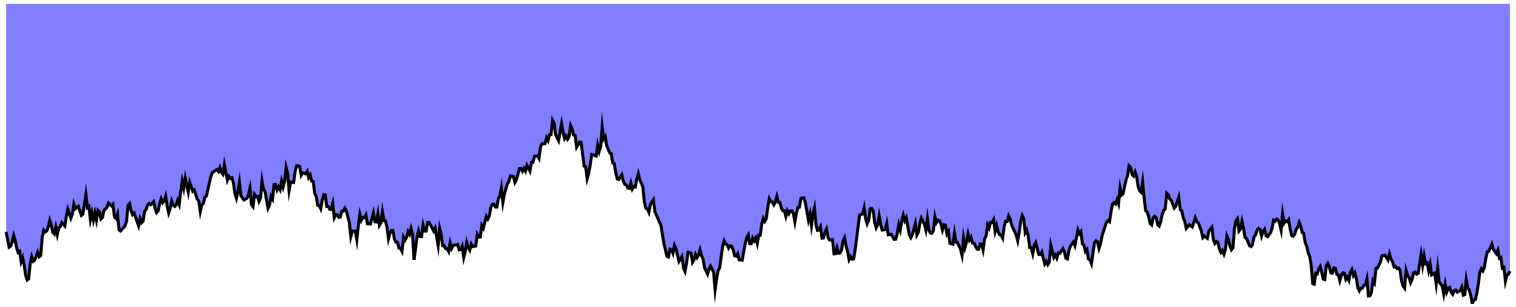
– the predictor: not well adapted as ℓ does not depend linearly of y

Alternative approach:

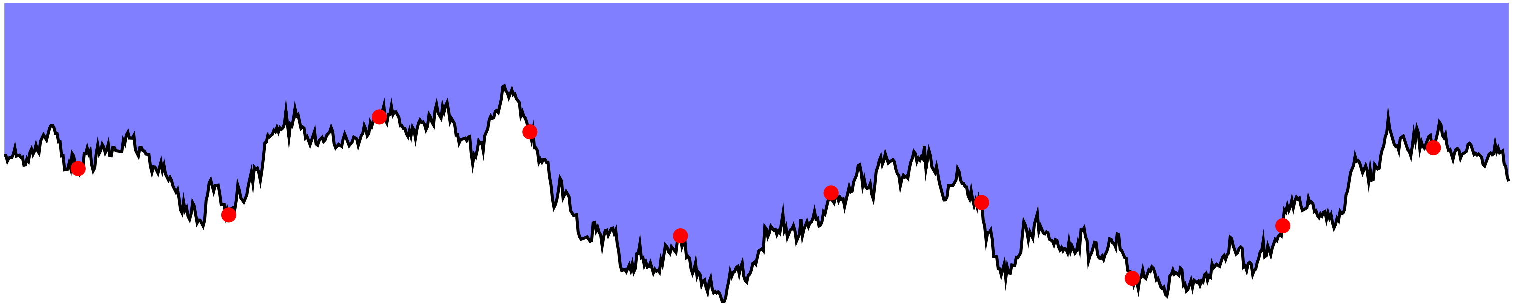
The sea-bed topography $\big(y(x), a \leq x \leq b\big)$ is considered as a realization of some stochastic process $Y = \big(Y(x), a \leq x \leq b\big)$, whose statistical features have to be chosen in full compatibility with the available data.

This makes it possible to predict ℓ by a Monte Carlo approach using conditional simulations.

Simulation and conditional simulation

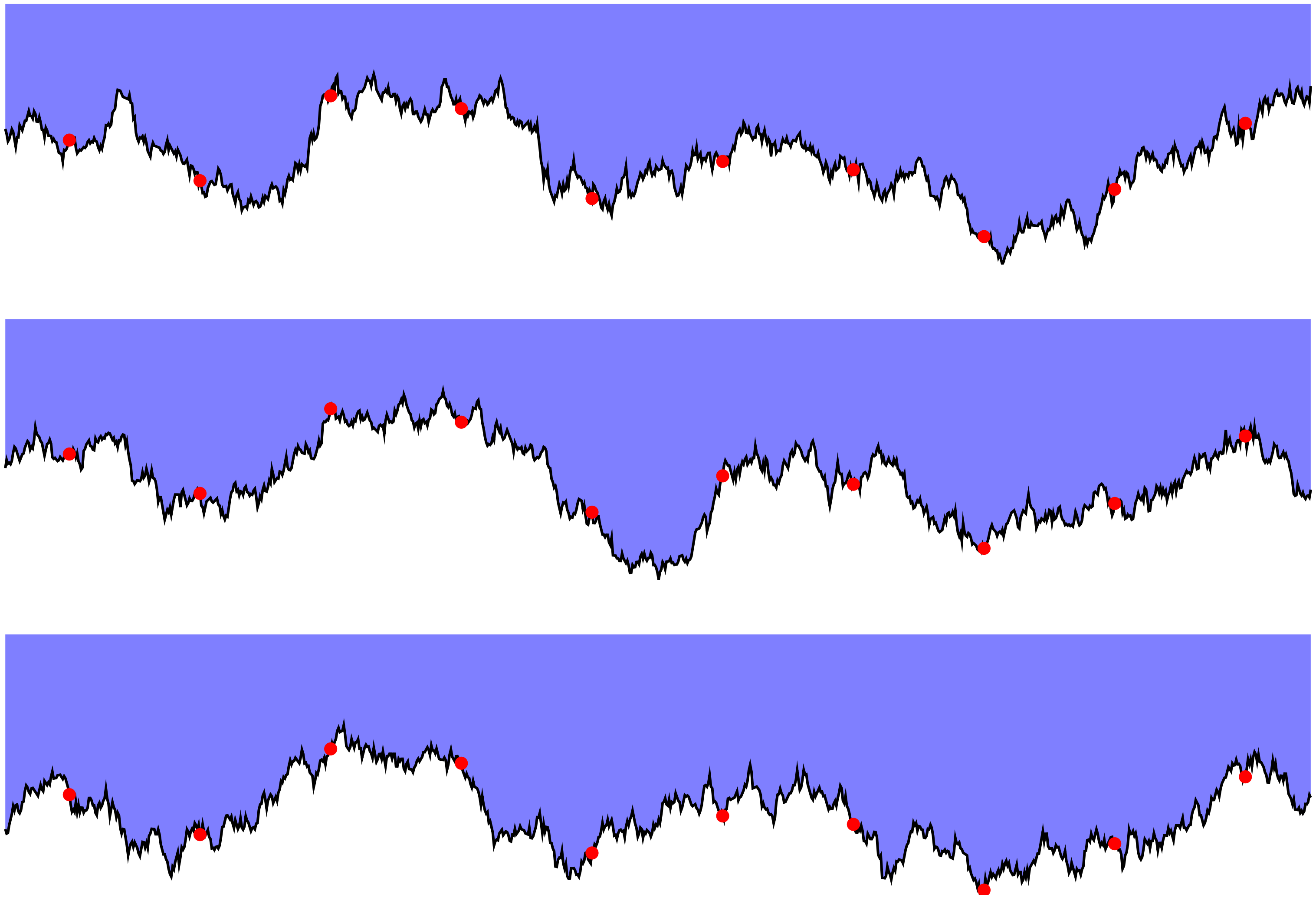


A simulation is a realization of the stochastic process.



A conditional simulation is a simulation that honors the available data.

Three examples of conditional simulations



Using conditional simulations

Suppose that a probabilistic model has been chosen. Given n conditional simulations $y_1, ..., y_n$, we calculate

$$\ell(y_i) = \int_a^b \sqrt{1 + [y_i'(x)]^2}\, dx \qquad i = 1, ..., n$$

from which we can

– predict the cable length (average of the $\ell(y_i)$'s);

– assign a precision for the predicted length (e.g. variance of the $\ell(y_i)$'s);

– give confidence limits for the predicted length (based on the histogram of the $\ell(y_i)$'s);

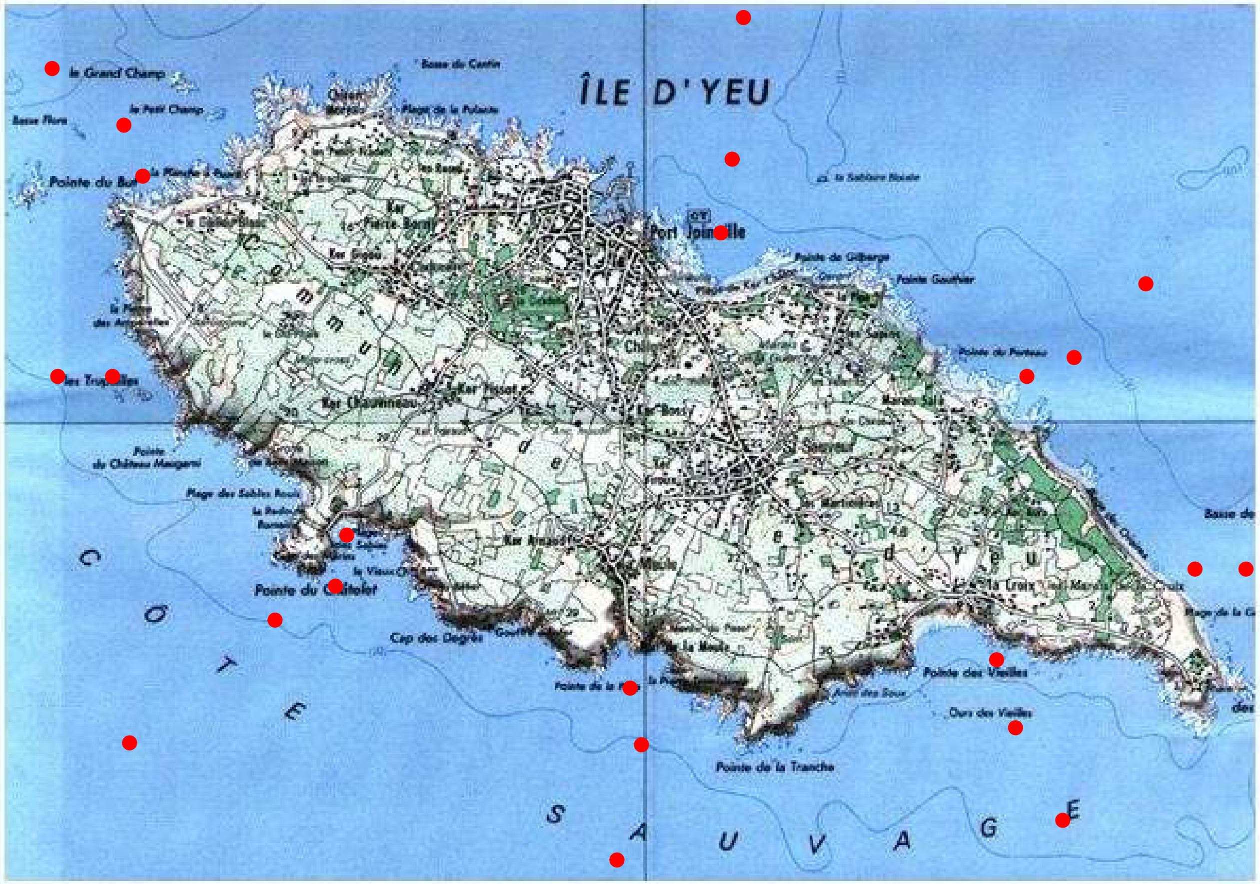
– predict the probability that the length does not exceed a given critical value.

The obtained results are totally tributary of the stochastic model that has been chosen.

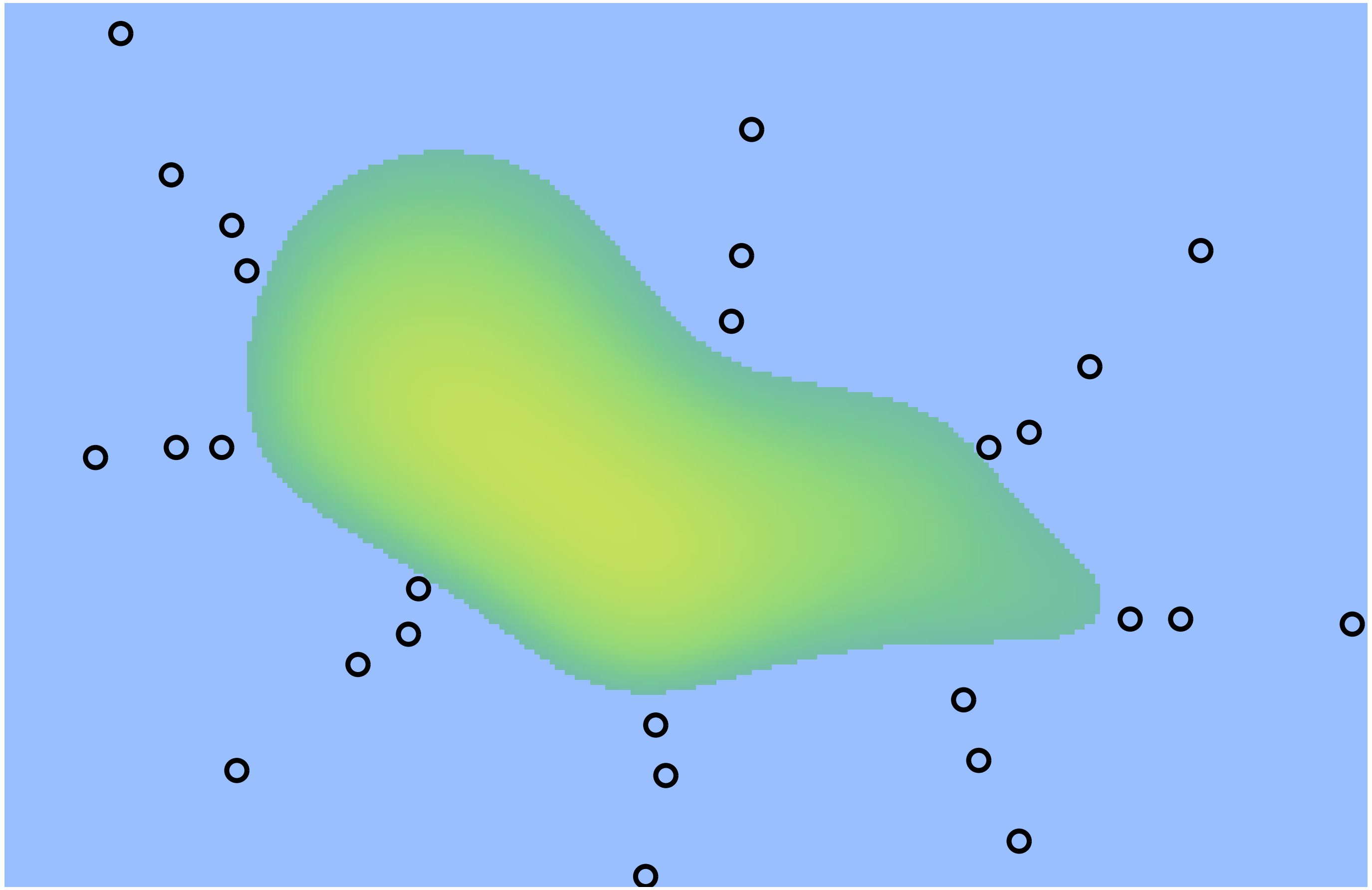
An illustration (after Chilès, 1977)

Yeu island

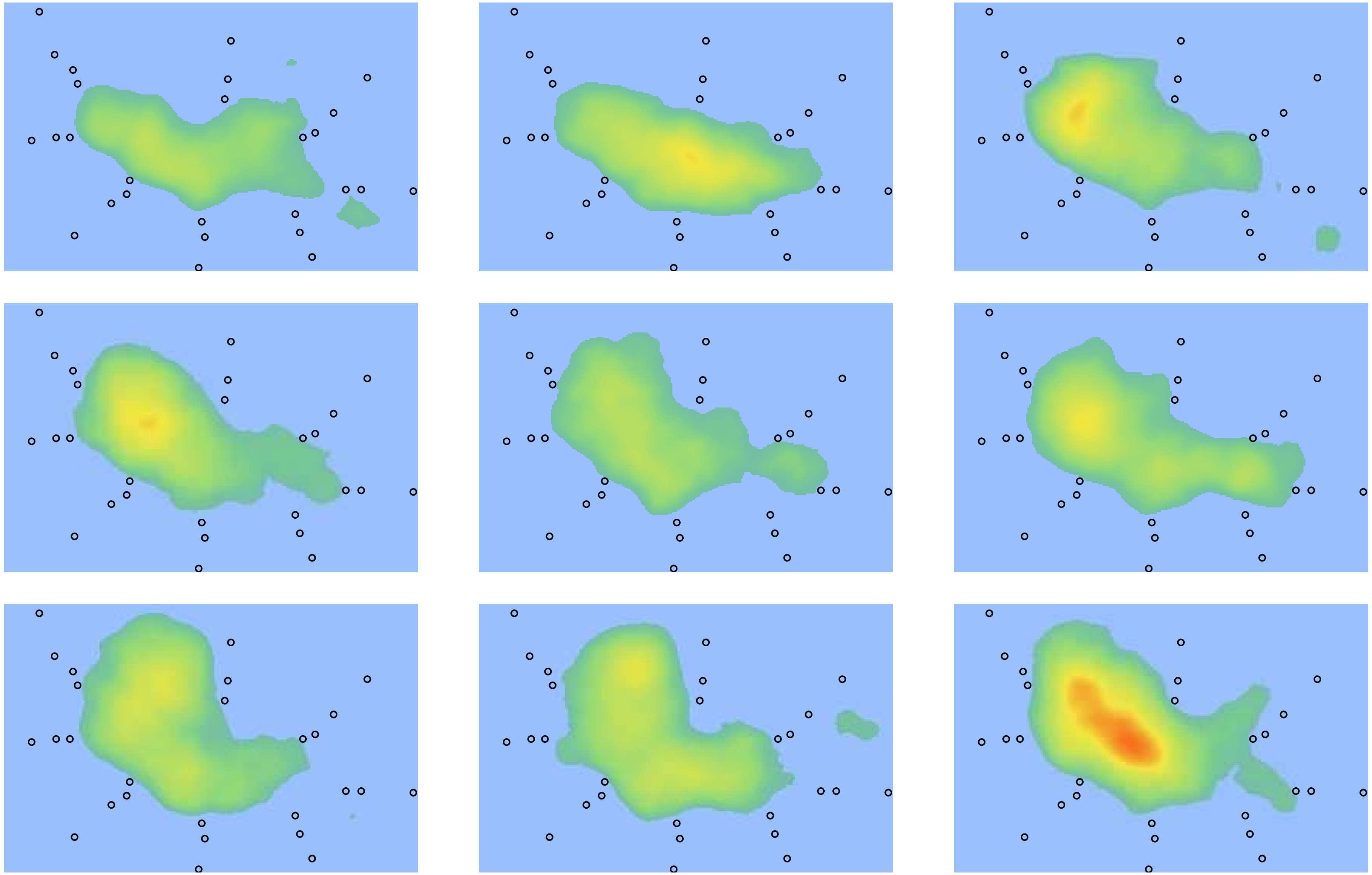
Samples


ÎLE D'YEU
Pointe du Châtelet
Cap des Degrés
Pointe des Vieilles
Pointe de la Tranche
C Ô T E
S A U V A G E

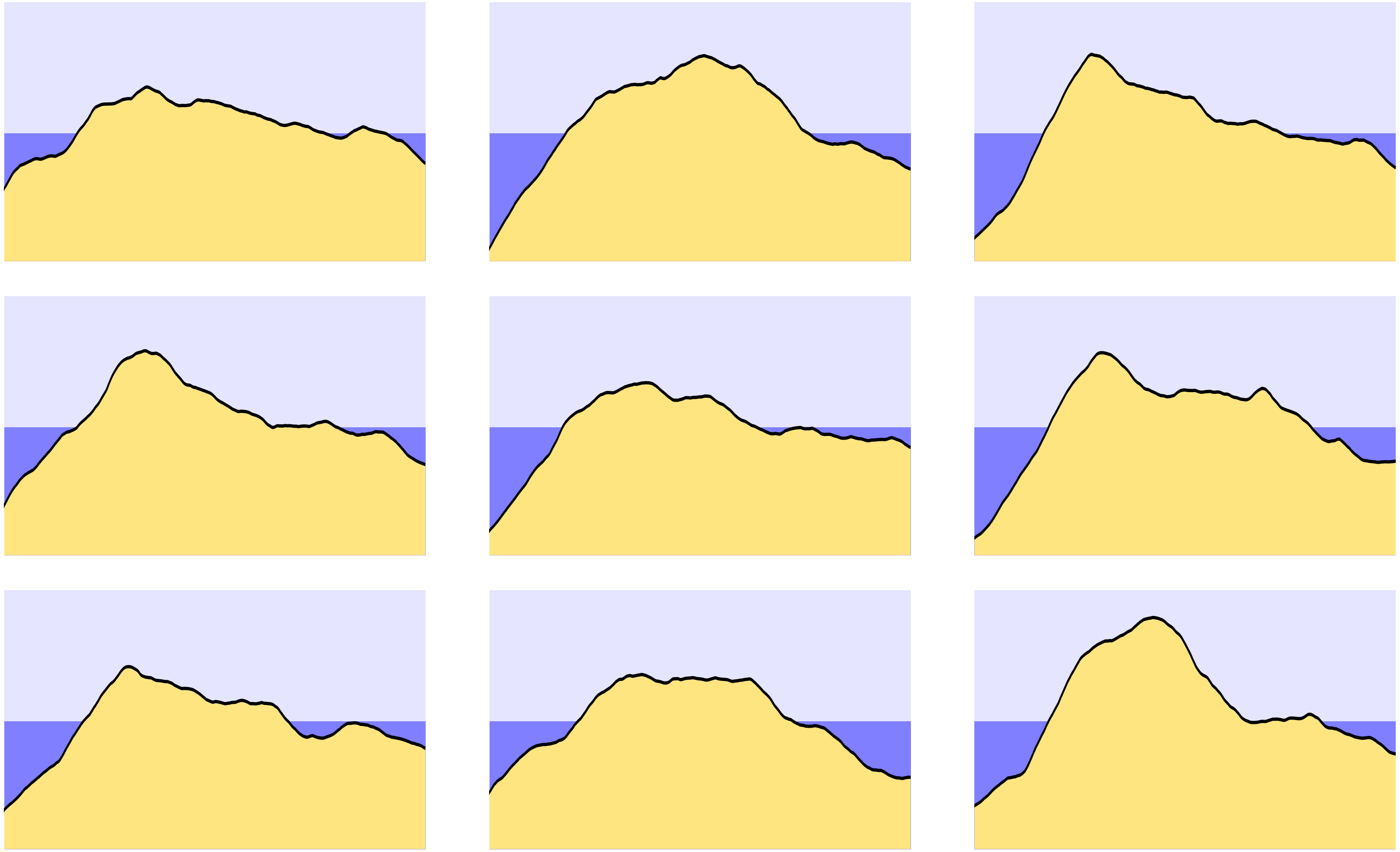
Prediction by linear regression (kriging)



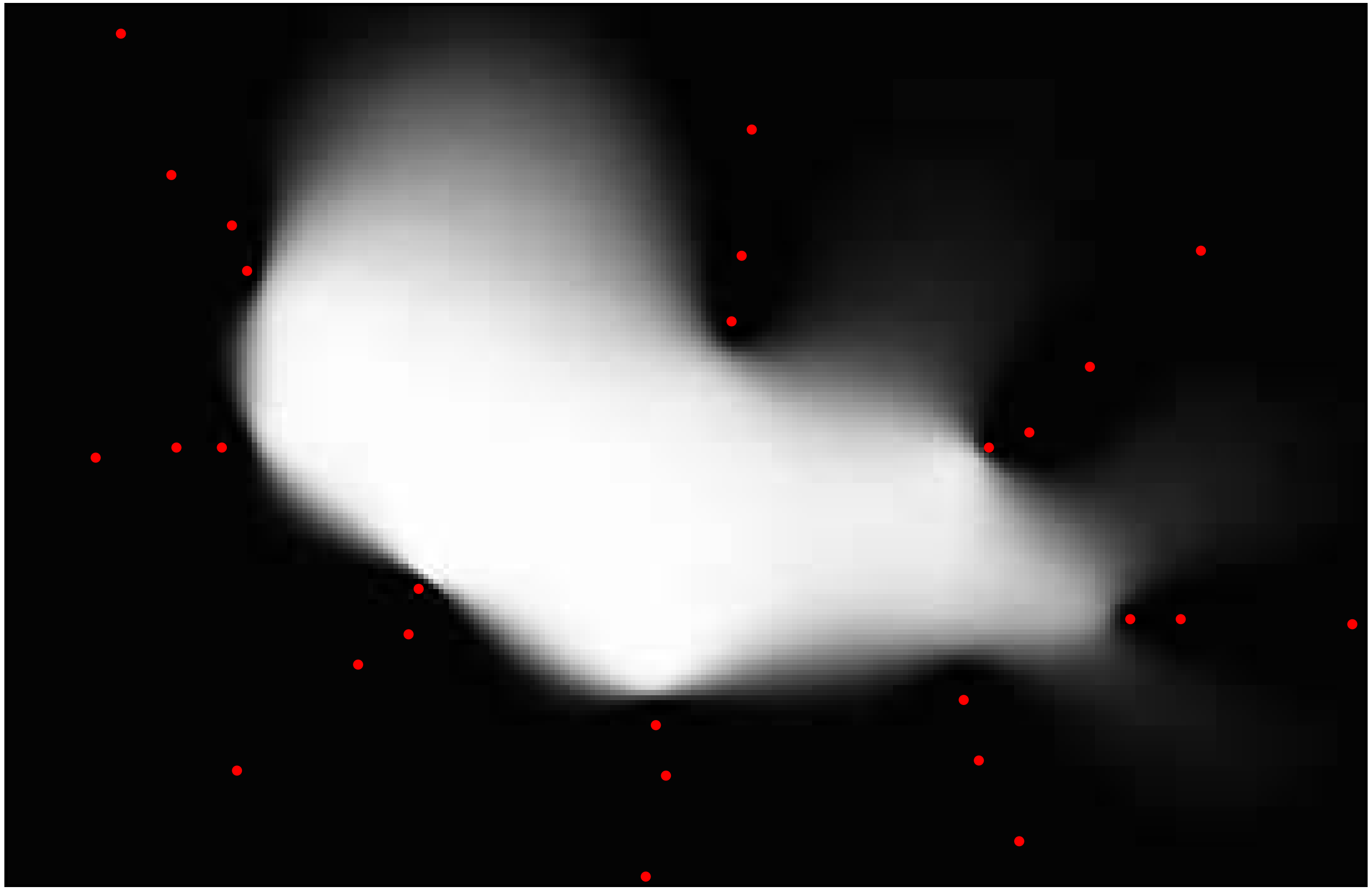
Conditional simulations



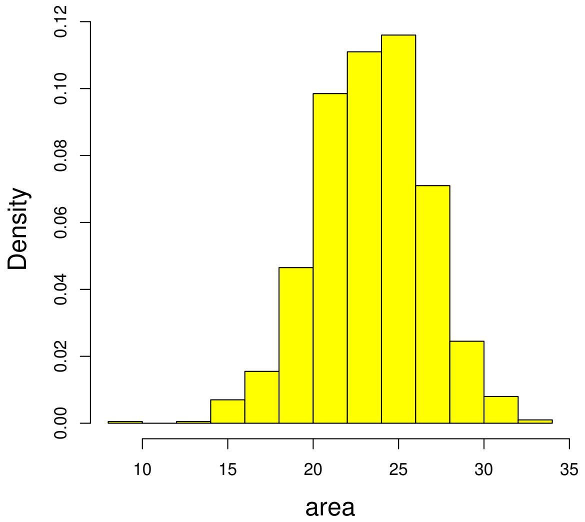
Simulated profiles

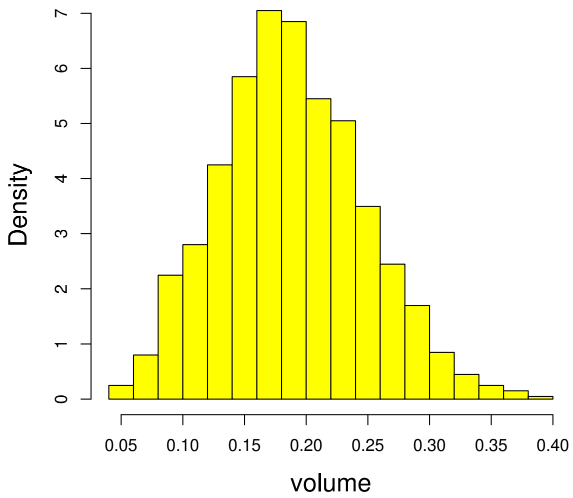


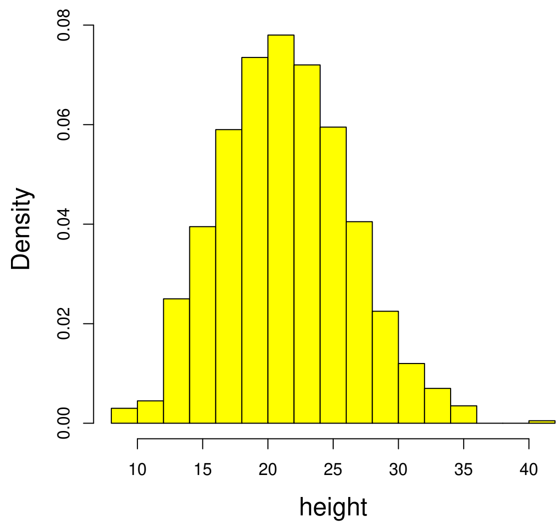
Probability that a point belongs to the island



Results

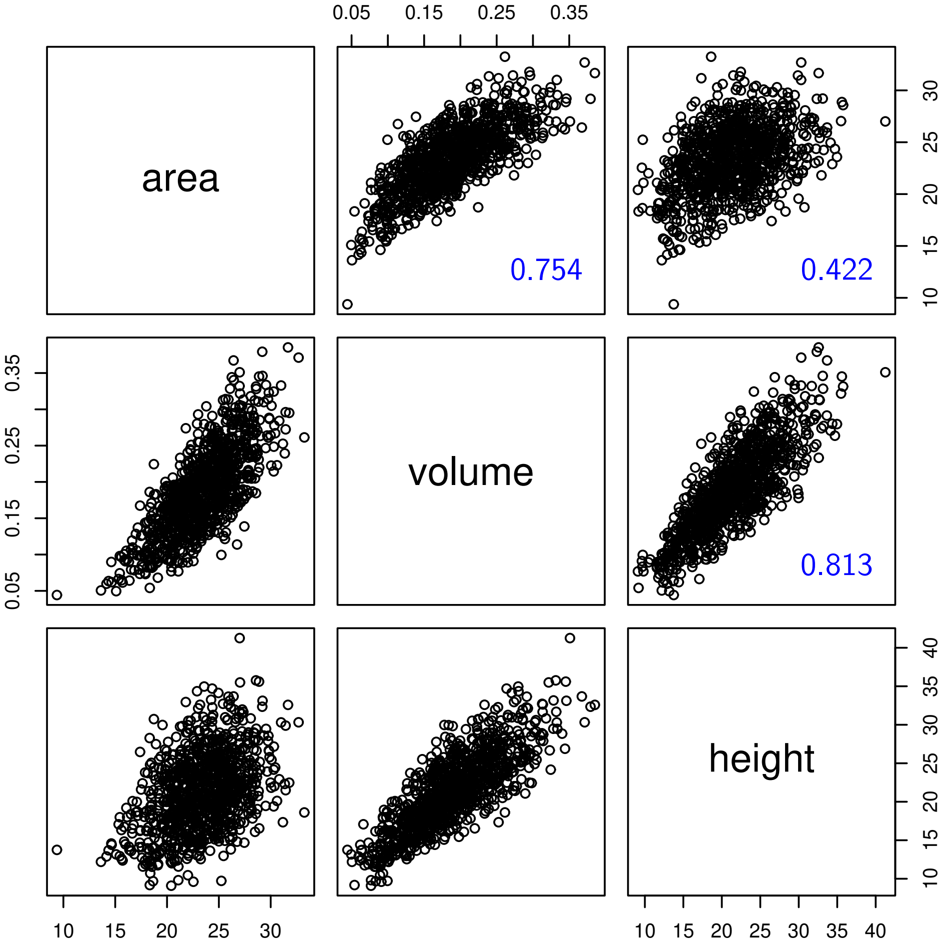




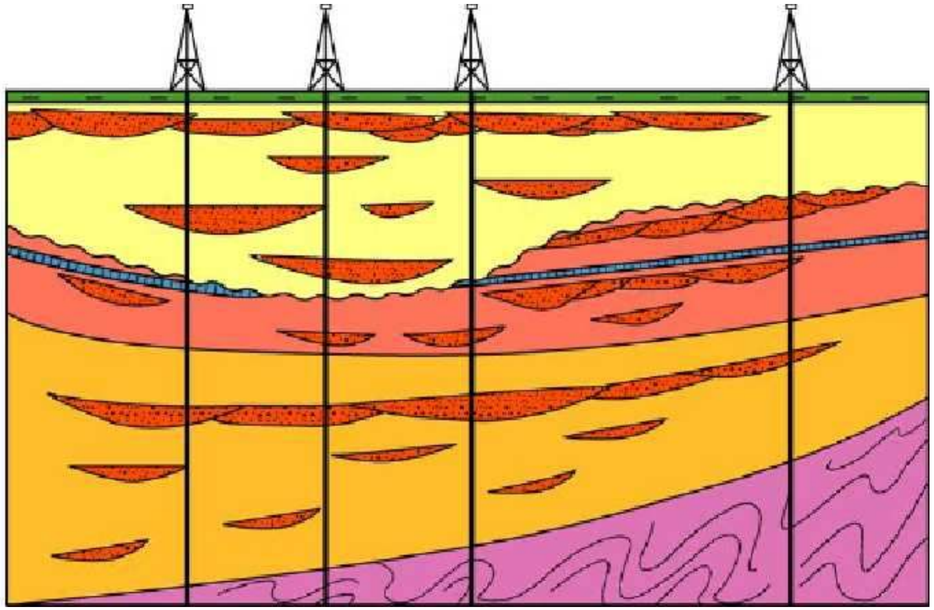


	Predicted	Simulated	Actual
Area (km^2)	22.94	23.37	23.32
Volume (km^3)	0.169	0.188	
Height (m)	15.93	21.32	27.50

Scatter diagrams



Motivation in reservoir engineering



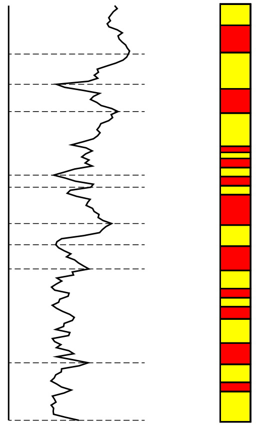
Case of a fluvial reservoir:

– the sandstone lenses or channels are porous and may contain oil;

– the clay facies acts as a barrier for the circulation of oil.

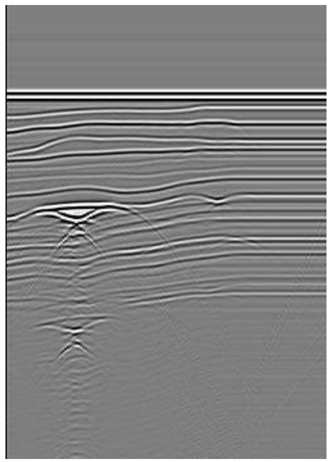
The geometry of the reservoir is required as an input of a fluid flow simulation exercise. How to predict it starting from the available data?

Data integration



– Diagraphic interpretation:

It provides facies information all along each well



– Sismic interpretation:

It provides the facies proportion within the reservoir

– Well tests:

They provide connectivity properties

Outline

Introductive example

Gaussian random field

– definition

– conditional simulation algorithm

Excursion set of a gaussian random function

– definition

– conditional simulation algorithm

Boolean model

– definition

– intersection property

– conditional simulation algorithm

Conclusions

Gaussian random field

Gaussian random field

Let $Y = (Y(x), x \in I\!R^d)$ be a 2^{nd} order stationary random field with mean m and covariance function C

Definition:

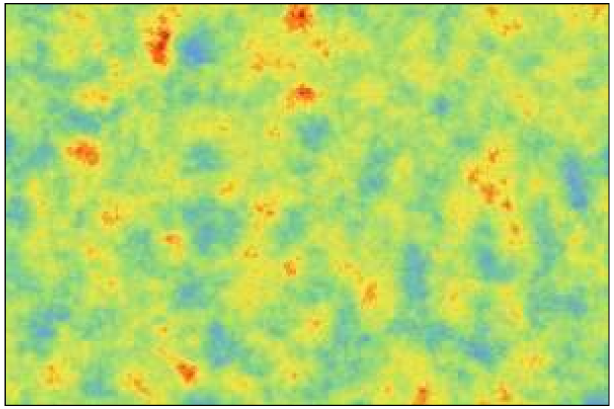
Y is said to be gaussian if any linear combination of its variables is gaussianly distributed:

$$\sum_{i=1}^{n} a_i Y(x_i) \;\sim\; \mathcal{G}\left(m \sum_{i=1}^{n} a_i, \sum_{i,j=1}^{n} a_i a_j C(x_i - x_j) \right)$$

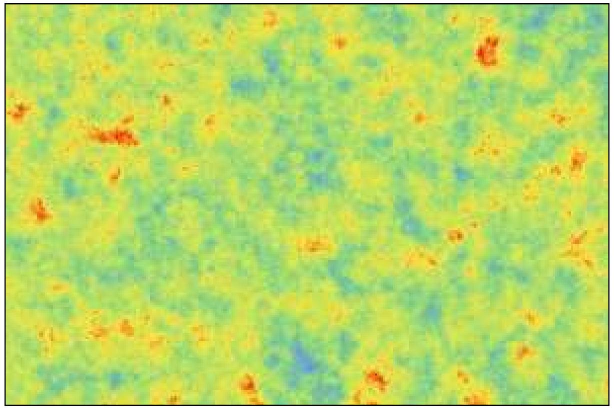
Fundamental property:

The spatial distribution of Y is totally characterized by its mean and its covariance function.

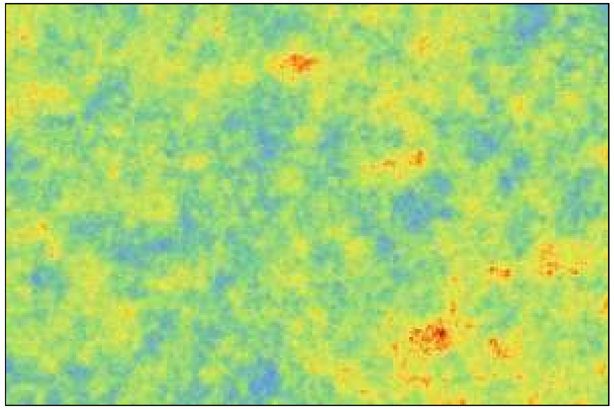
Examples (different covariance functions)



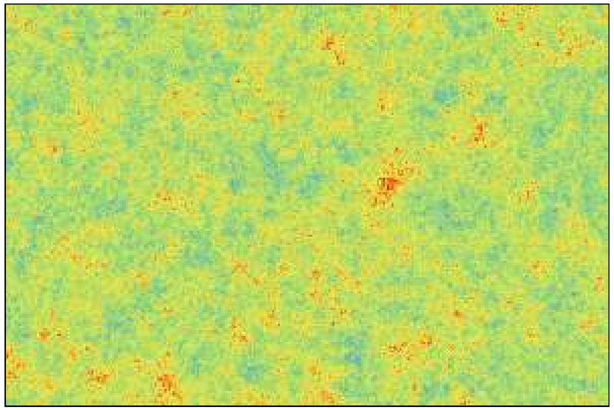
spherical



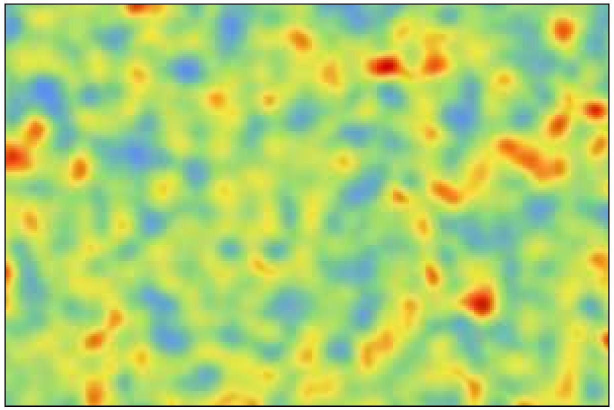
exponential



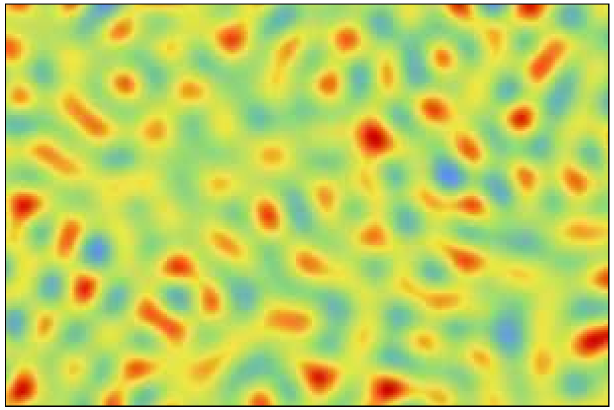
hyperbolic



stable



gaussian

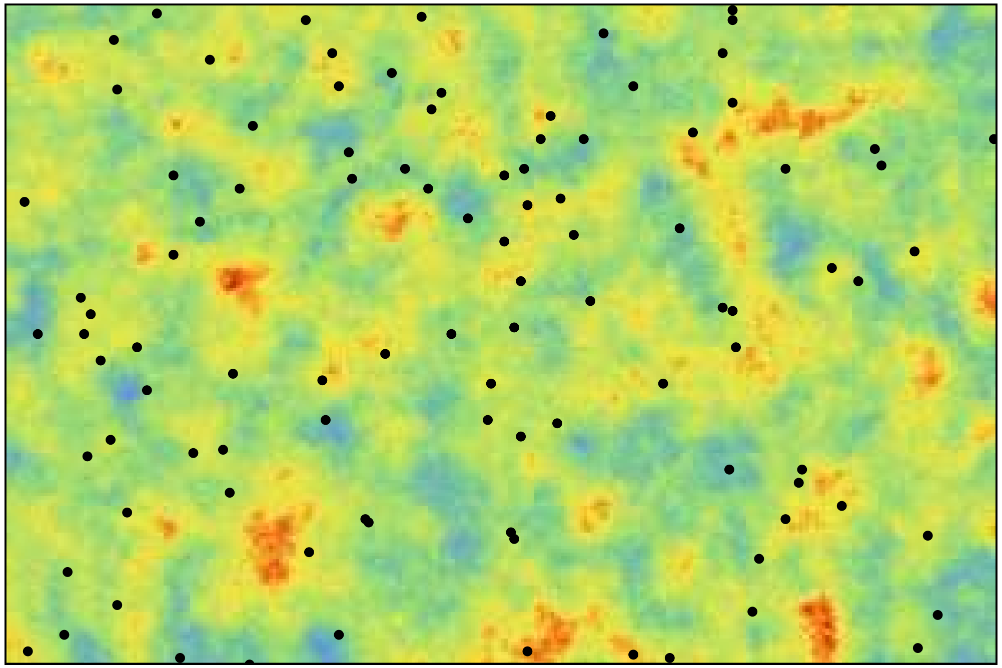


cardinal sine

Presentation of the problem

How can we produce a realisation y of a stationary gaussian random field with mean 0, covariance function C, that satisfies

$$y(x_\alpha) = y_\alpha \qquad \alpha \in A$$



Principle

Write

$$Y(x) = Y^A(x) + Y(x) - Y^A(x) \qquad x \in \mathbb{R}^d$$

where

$$Y^A(x) = \sum_{\alpha \in A} \lambda_\alpha(x) Y(x_\alpha) \qquad \text{kriging estimator (known)}$$

$$Y(x) - Y^A(x) \qquad \text{kriging residual (unknown)}$$

Y^A and $Y - Y^A$ are two independent gaussian random fields.

Accordingly, put

$$Y^{CS}(x) = Y^A(x) + S(x) - S^A(x) \qquad x \in \mathbb{R}^d$$

where S is an independent copy of Y. $S - S^A$ is known insofar as realizations of S can be produced (non conditional simulations).

Conditional simulation
Algorithm and verifications

Algorithm:

(i) generate a simulation s. Put $s_\alpha = s(x_\alpha)$ *for each* $\alpha \in A$*;*

(ii) for each $x \in \mathbb{R}^d$*, do*

(ii.i) compute the kriging weights $\big(\lambda_\alpha(x), \alpha \in A\big)$*;*

(ii.ii) return $y^{CS}(x) = y^A(x) + s(x) - s^A(x) = s(x) + \sum_\alpha \lambda_\alpha(x)(y_\alpha - s_\alpha)$.

Verifications:

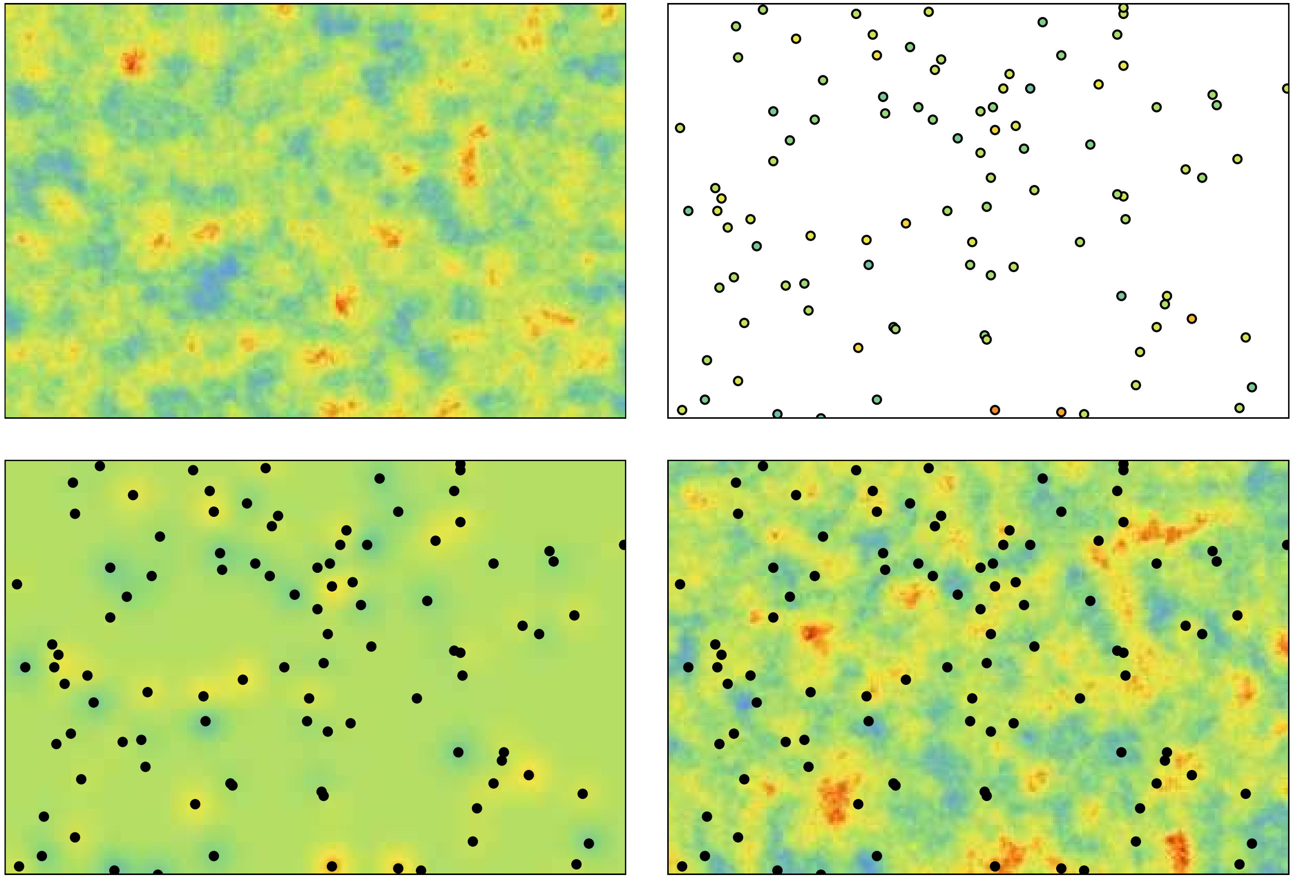
– if $x = x_\alpha$, then $y^{CS}(x_\alpha) = y_\alpha + s(x_\alpha) - s_\alpha = y_\alpha$

(conditioning data are honored)

– if $C(x - x_\alpha) \approx 0$ for each $\alpha \in A$, then $y^{CS}(x) \approx 0 + s(x) - 0 = s(x)$

(far from the data points, the conditional simulation is non conditional)

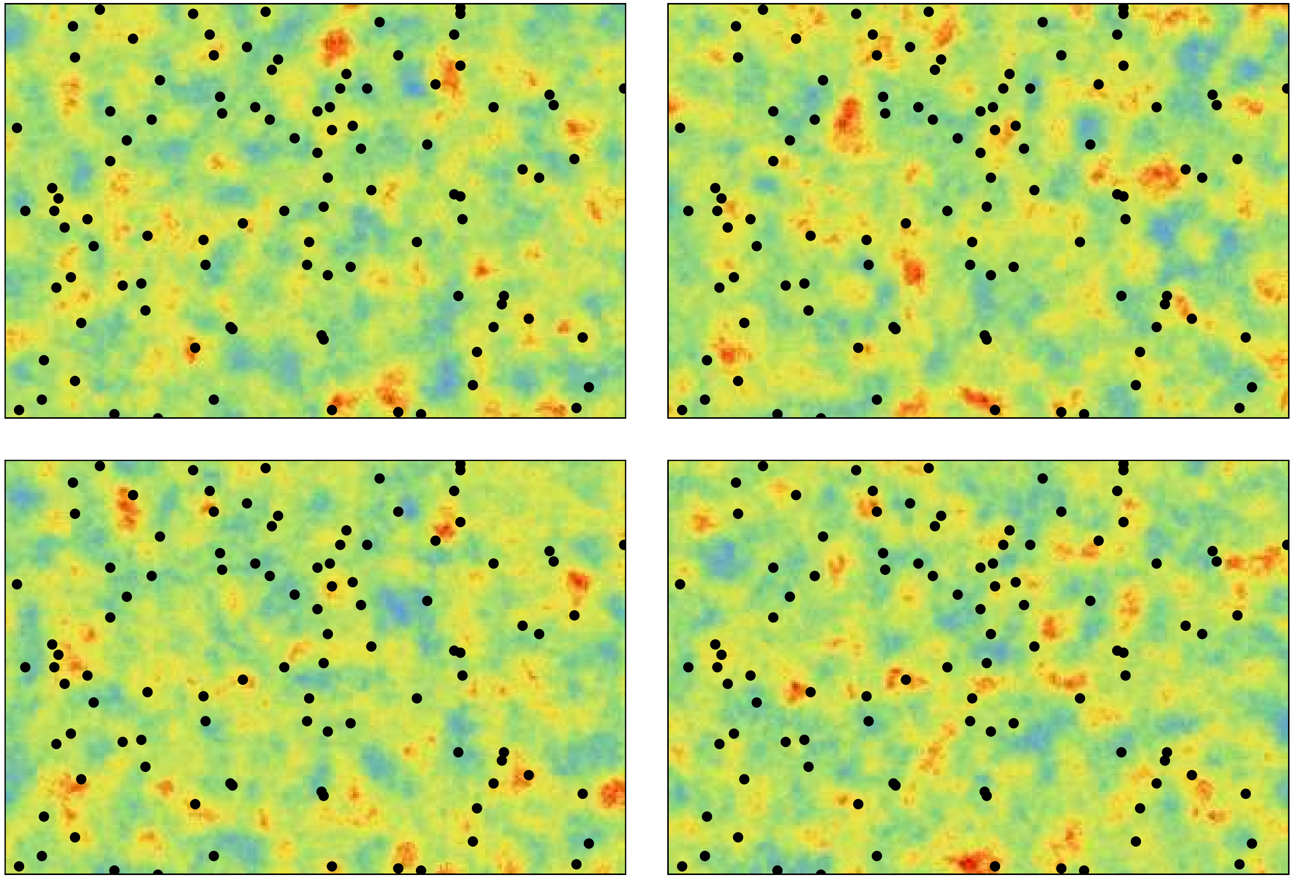
Illustration



Model $m = 0$, $C = sph(1, 20)$. Simulation field 300×200.

Simulation (TL), conditioning data points (TR), kriging estimate (BL) and conditionnal simulation (BR).

Four conditional simulations



Excursion set
of a gaussian random field

Excursion set of a gaussian random field

Basic ingredients:

– a 2^{nd} order stationary, standardized gaussian random field Y with covariance function C.

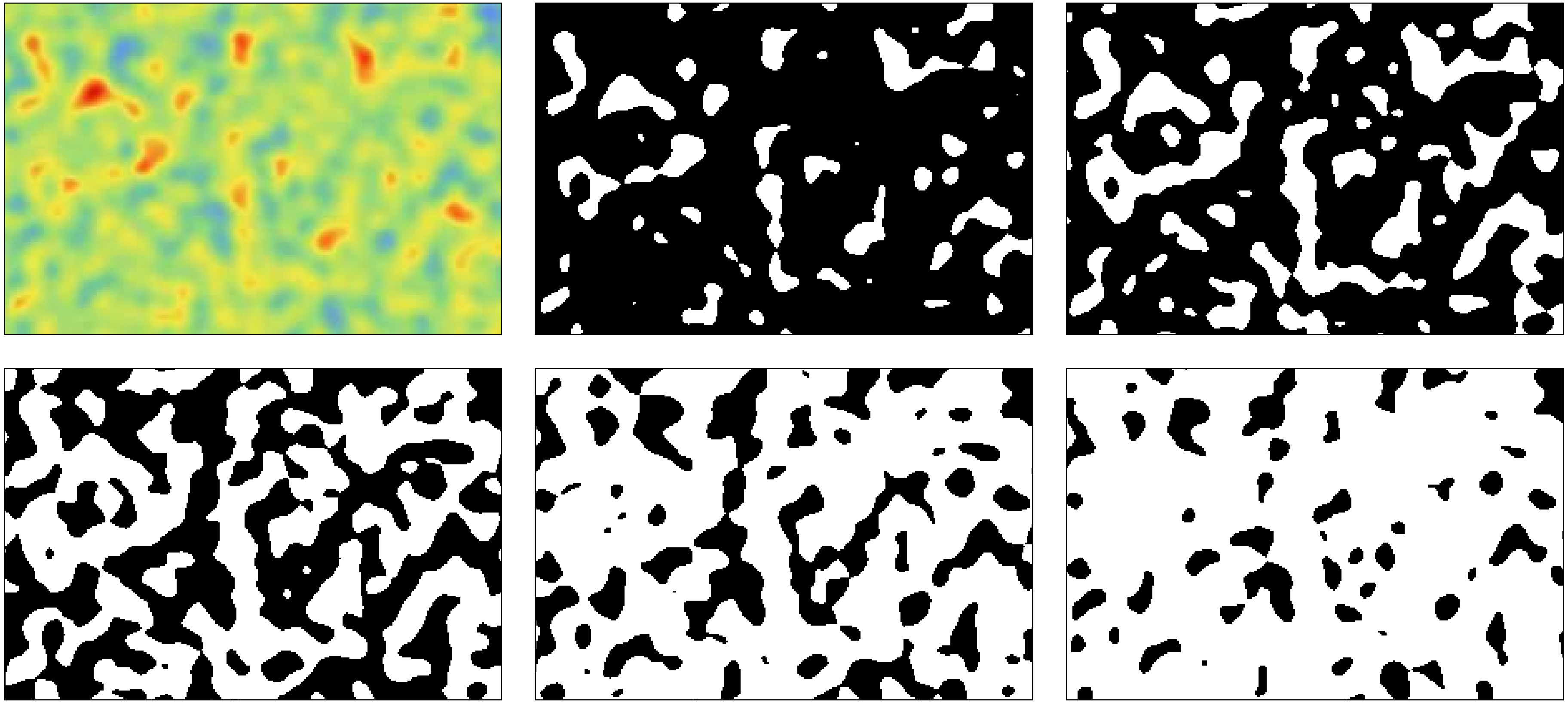
– a numerical value λ.

Definition:

The excursion set of Y at level λ is the set of points where Y takes values greater or equal to λ

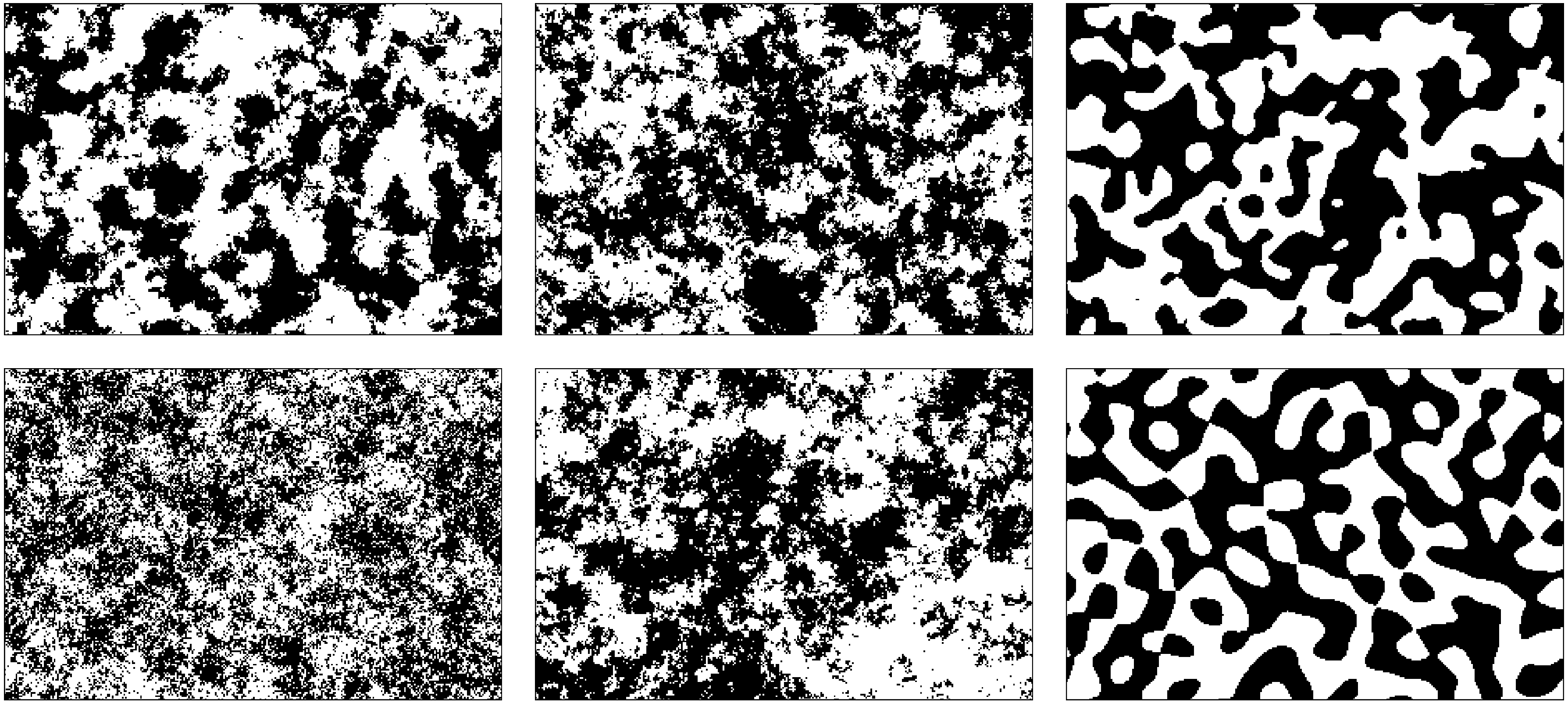
$$X_\lambda(x) = \begin{cases} 1 & \text{if } Y(x) \geq \lambda \\ 0 & \text{if } Y(x) < \lambda \end{cases}$$

Examples at various levels



Gaussian random field (gaussian covariance) and its excursion sets at levels $-1, -0.5, 0, 0.5$ et 1.

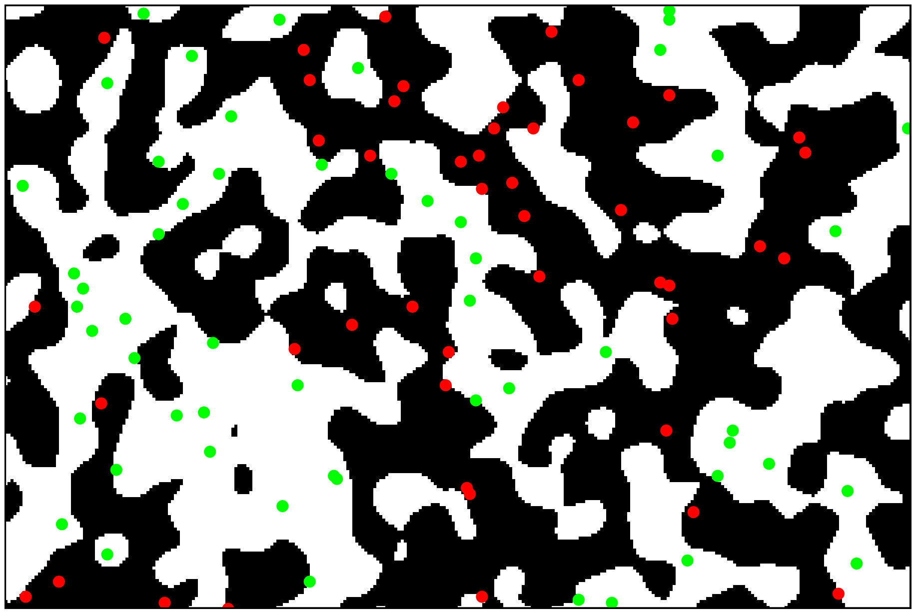
Examples for different covariance functions



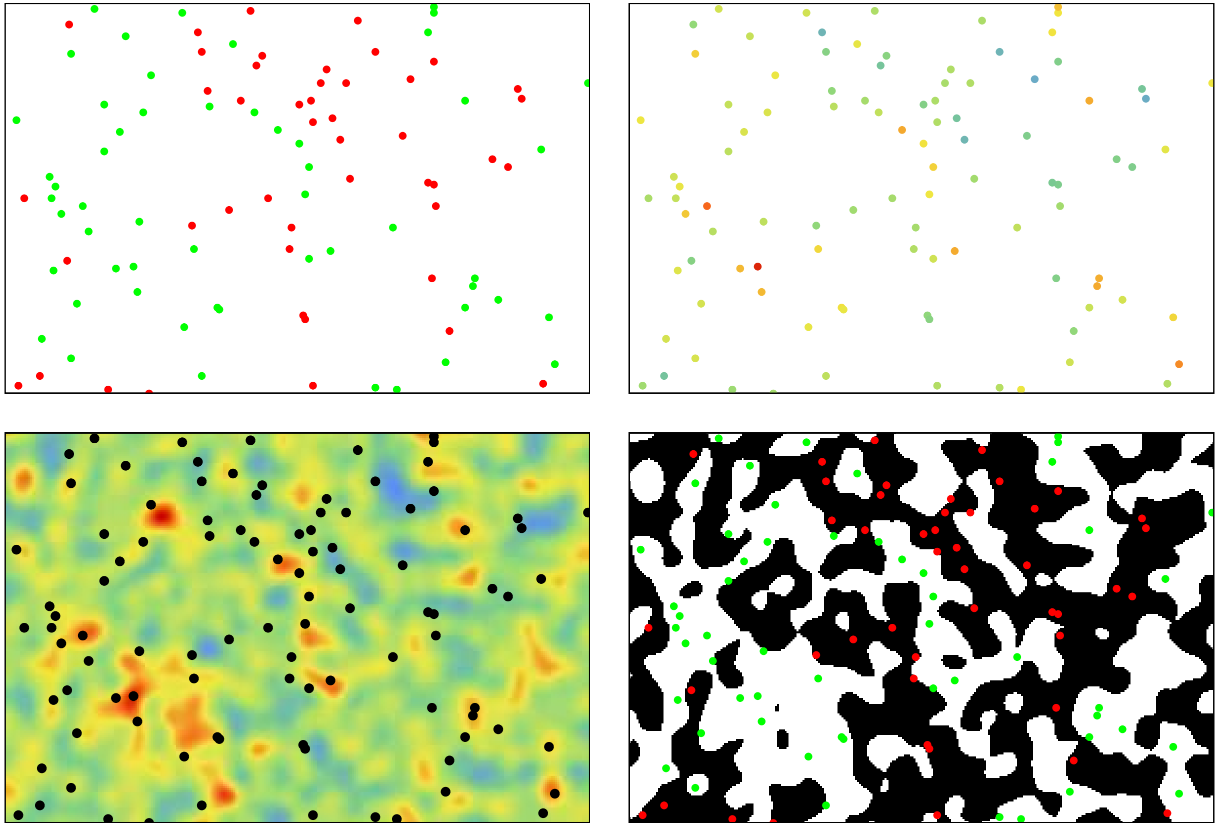
Excursion sets at level 0 derived from gaussian random fields with spherical (TL), exponential (TM), gaussian (TR), stable (BL), hyperbolic (BM) and cardinal sine (BR) covariance functions.

Conditional simulation
Presentation of the problem

How to produce realizations of the excursion set at level λ of a standardized gaussian random field with covariance function C, in such a way that each facies contains a finite number of prespecified points?

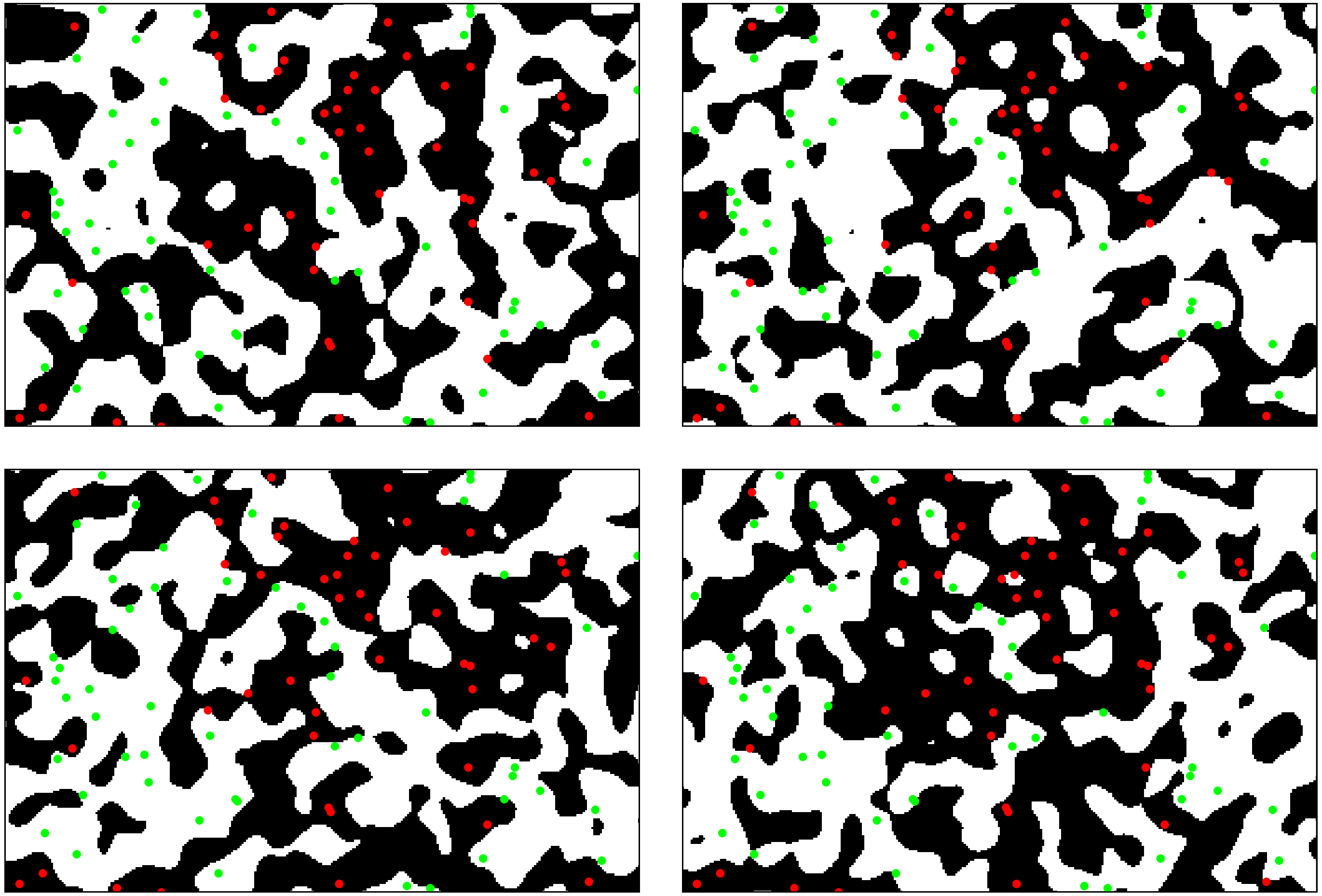


Conditional simulation algorithm



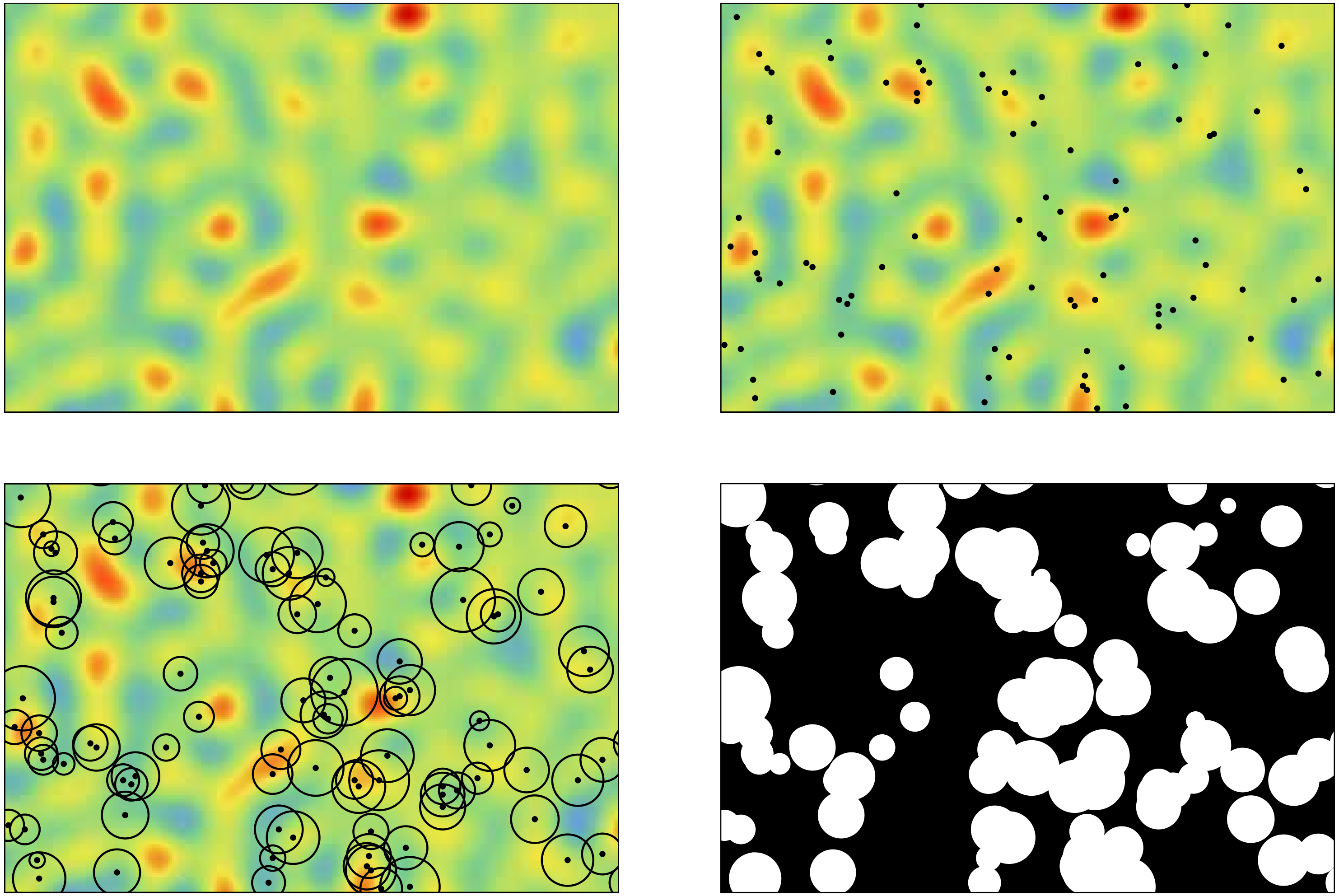
Conditioning data set (TL). Conditional simulation of Y at the data points only (Gibbs sampler) (TR). Conditional simulation of Y (BL). Threshold at level λ (BR).

Four conditional simulations



Boolean model

Constructing a Boolean model



Definition of a boolean model

Basic ingredients:

– a Poisson point process $\mathcal{P}$ with intensity function $\theta = \big(\theta(x), x \in I\!R^d\big)$;

– a family $\big(A(x), x \in I\!R^d\big)$ of independent nonempty compact random subsets (referred to as "objects"). The statistical properties of $A(x)$ can be specified by its distribution function

$$T_x(K) = P\{A(x) \cap K \neq \emptyset\} \qquad K \in \mathcal{K}(I\!R^d)$$

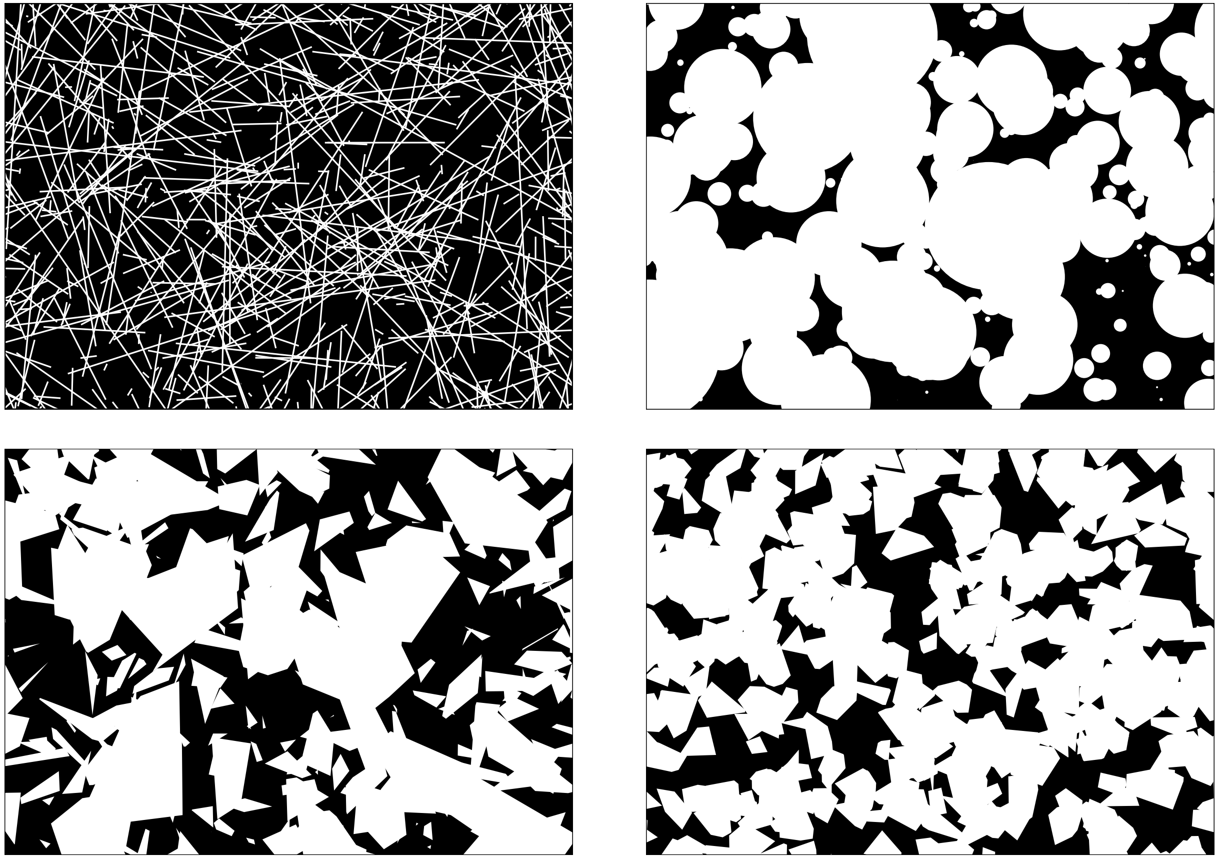
Definition:

A Boolean model is the union of the objects located at the Poisson points.

$$X = \bigcup_{x \in \mathcal{P}} A(x)$$

$X =$ foreground $\qquad X^c =$ background

Examples of boolean models



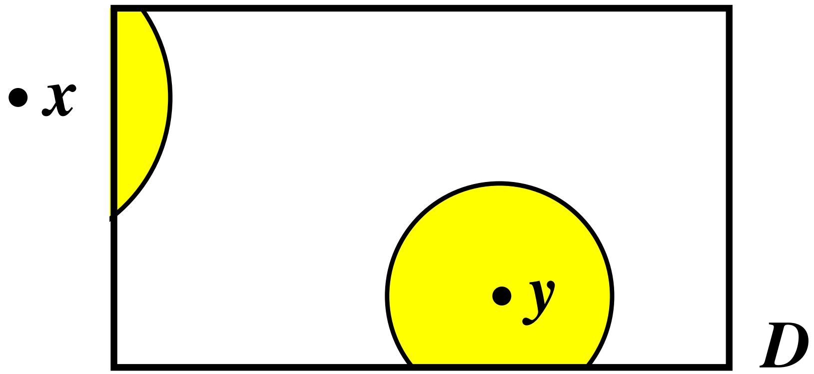
Boolean models of segments (TL), disks (TR), Poisson (BL) and Voronoi polygons (BR).

Intersection of the model and a domain

If X is a Boolean model with parameters (θ, T) and D is a compact domain, then $X \cap D$ is also a boolean model with parameters $(\theta^{(D)}, T^{(D)})$ given by

– Poisson intensity $\qquad \theta^{(D)}(x) = \theta(x) T_x(D) \qquad x \in \mathbb{R}^d$

– distribution function $\qquad T_x^{(D)}(K) = \dfrac{T_x(K)}{T_x(D)} \qquad K \in \mathcal{K}(D)$



Typical objets

Assumption:

The Poisson intensity $\theta^{(D)}$ has a finite integral $\vartheta(D)$.

Consequences:

– $X \cap D$ has a finite (Poisson) number of objects;

– $\theta^{(D)}(\cdot)/\vartheta(D)$ is a probability density function on $\mathbb{R}^d$.

Definition:

The random compact set $A(\dot{x}) \mid A(\dot{x}) \cap D \neq \emptyset$ with $\dot{x} \sim \theta^{(D)}/\vartheta(D)$ is called a typical object (hitting D). Its distribution function is (for $K \subset D$)

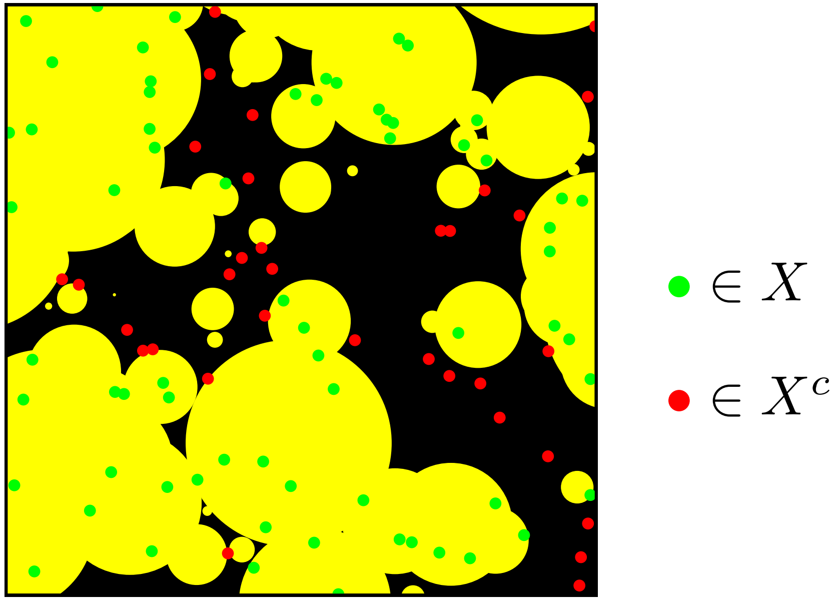
$$T^{(D)}(K) = \int_{\mathbb{R}^d} \frac{\theta^{(D)}(x)}{\vartheta(D)} \frac{T_x(K)}{T_x(D)} \, dx = \frac{\int_{\mathbb{R}^d} \theta(x) \, T_x(K) \, dx}{\int_{\mathbb{R}^d} \theta(x) \, T_x(D) \, dx}$$

Property:

$X \cap D$ is the union of a Poisson number (mean $\vartheta(D)$) of independent typical objects.

Conditional simulation
Presentation of the problem

How to produce realizations of X in the domain D subject to the conditions that one finite set of points C_0 must be contained in X^c and another set C_1 in X?

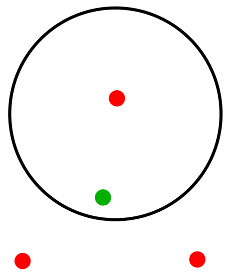


Compatibility between model and data

In contrast to a gaussian random function, a boolean model may not be compatible to a data set.

Example:

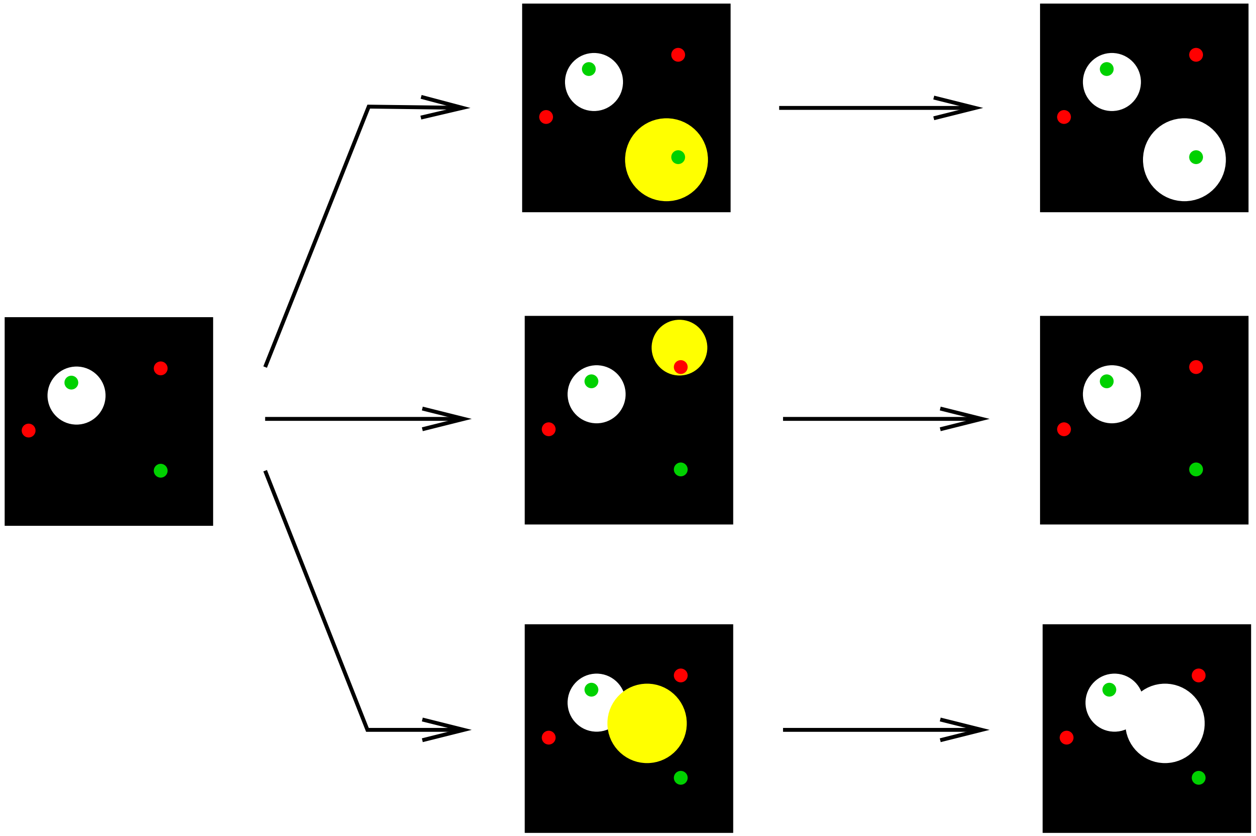
Objects are circular with fixed radius, and a foreground conditioning point is closely surrounded by background conditioning points:



We have not succeeded in finding a "direct" algorithm for simulating a boolean model conditionally.

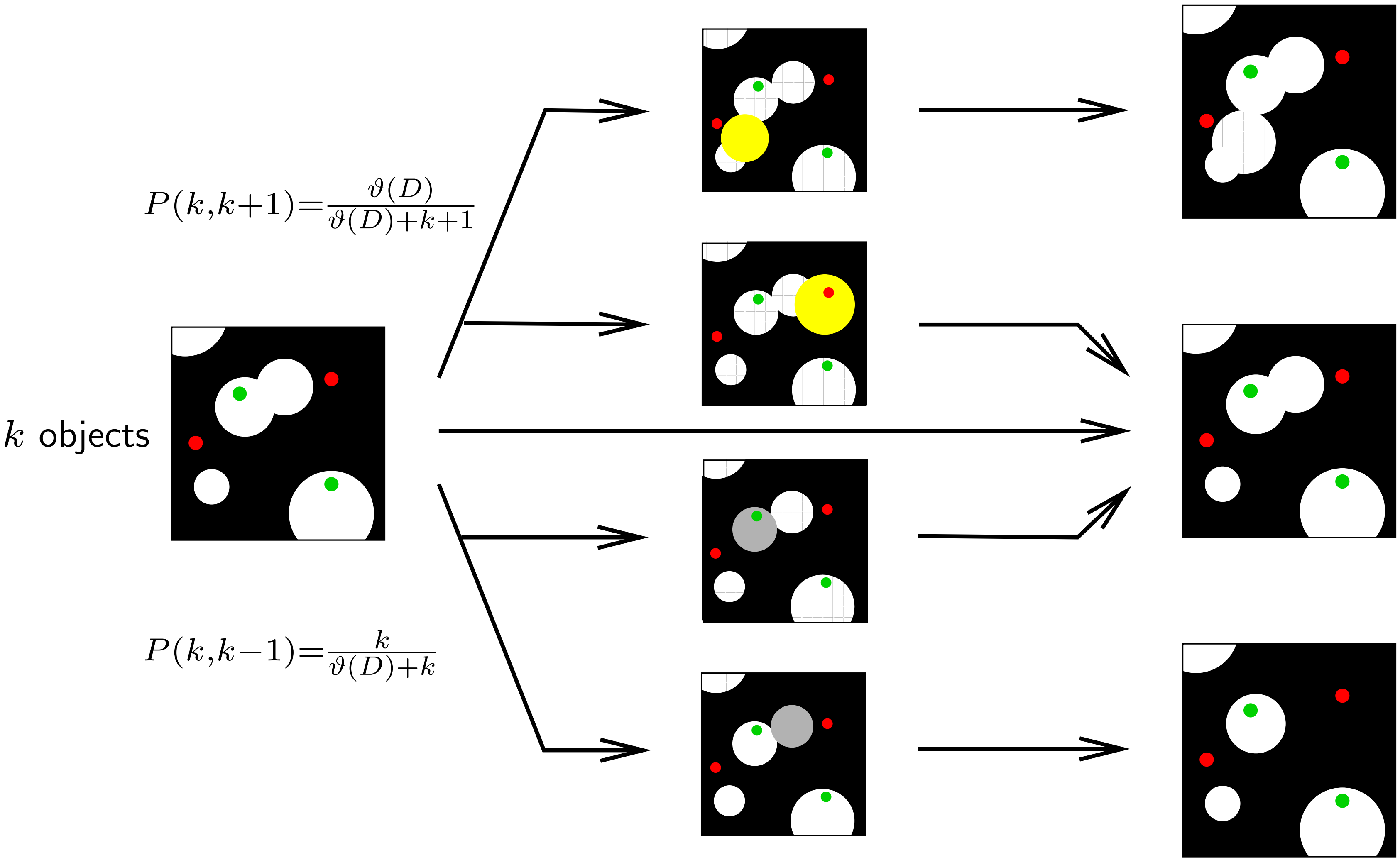
In what follows, an iterative algorithm is presented.

Initialization

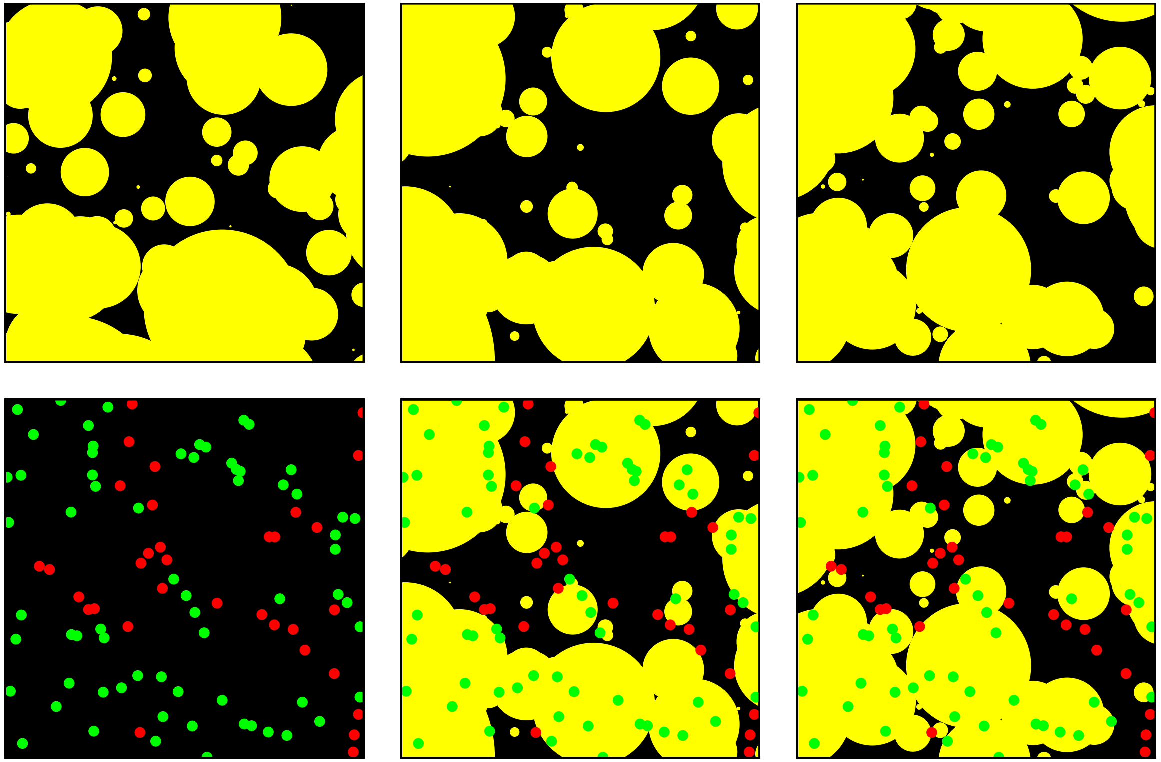


Principle: independent typical objects are sequentially generated. Any object that covers a conditioning point of the background is rejected. The initialisation procedure stops when all foreground conditioning points have been covered by objects.

One step of conditional simulation

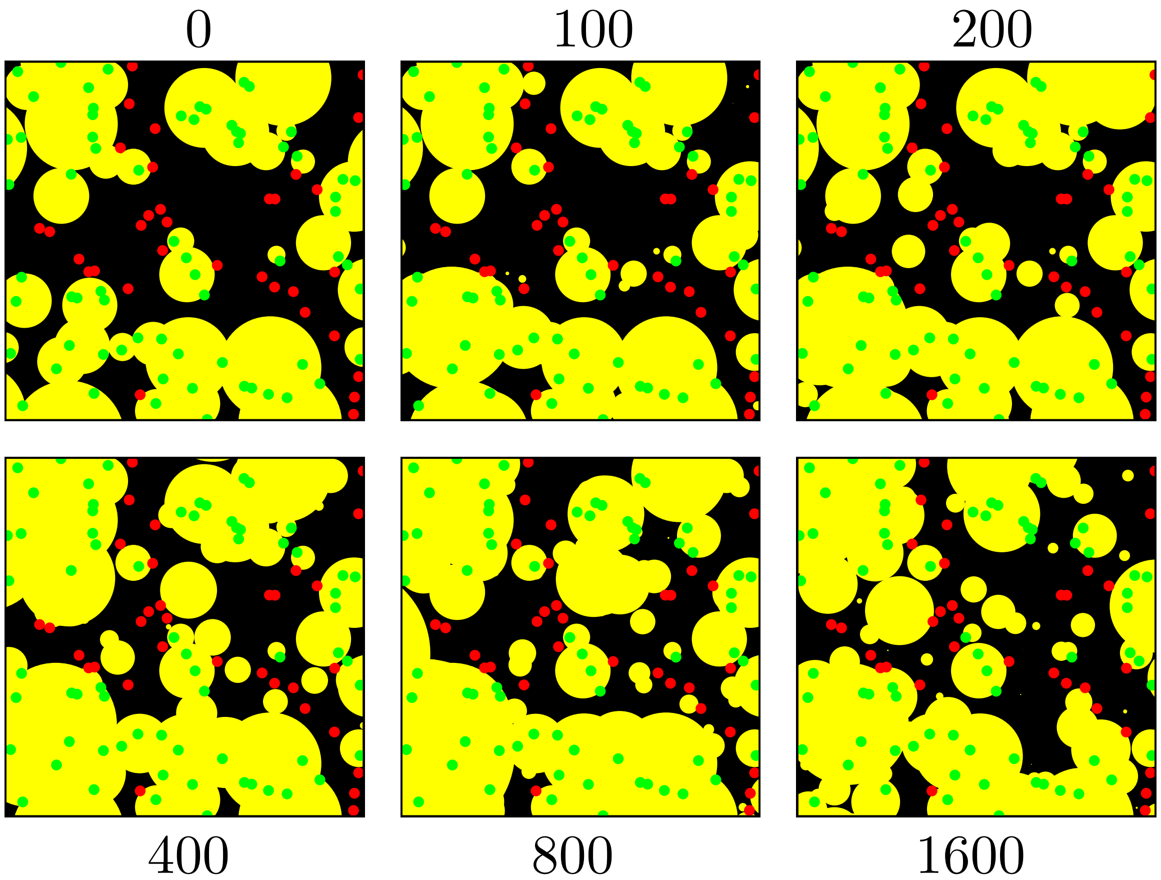


Example of conditional simulation



Stationary boolean model of disks (Poisson intensity 0.00766; radii are exponentially distributed with mean 5). The simulation field is 100×100. Left, a non conditional simulation and 100 conditioning data points. Middle and right, two conditional simulations.

Display of the simulation at various steps



Display of the conditional simulation at steps $0, 100, 200, 400, 800$ and 1600.

Conclusions

For a number of prediction problems, conditional simulations are an interesting possibility in numerous fields of the earth sciences:

– They automatically provide fully compatible estimates for all features of interest;

– They can also assign each of them a precision (variance or confidence limits);

Numerous difficulties remain

– Model choice: The obtained results are totally dependent on the stochastic model that has been chosen;

– Algorithm: design, rate of convergence (perfect simulations?)

– Data integration: accounting for complex data (non-linearity, different supports...)