Mini-Course on *Advanced Stationary Processes Analysis*, VIASM. Part 2: Geostatistics *Chapter 1: Probability Models*

Jean-Yves Dauxois

INSA - Toulouse Mathematical Institute, Toulouse, FRANCE
jean-yves.dauxois@insa-toulouse.fr

18th to 22th July 2016

Let

- $S \subset \mathbb{R}^d$ be a spatial set
- $(\Omega, \mathcal{F}, P)$ be a probability space
- $(E, \mathcal{E})$ be a measurable set.

Definition

A random field X, also called a spatial process, is a family $X = \{X_s, s \in S\}$ of random variables (r.v.), indexed by $s \in S$, from $(\Omega, \mathcal{F}, P)$ to $(E, \mathcal{E})$.

- S =spatial set of sites
- E =state space of the process.

The distribution P of X is unknown in the family $\mathcal{P}$ of probabilities on the measurable space $(E, \mathcal{E})$.

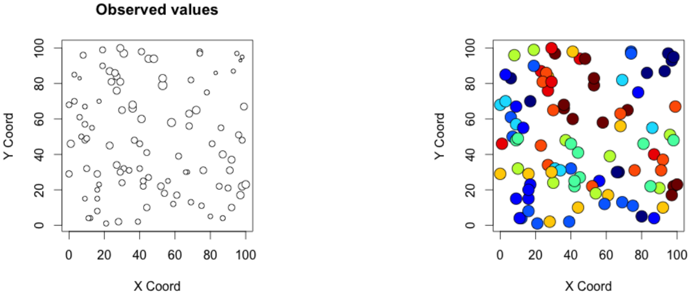

Figure: Observations of a Spatial Process on the square $[0, 100]^2$

Some problems to solve:

- Forecast X_{s_0} when the random field X is not observed at s_0
- Estimate the distribution of X_s or $\varphi(X_s)$
- Estimate the dependency between the X_{s_i}.

Definition

A spatial process $X = \{X_s, s \in S\}$ *is said of* **second order** *if, for all* s *in* S*, we have:*

$$\mathbb{E}X_s^2 < +\infty.$$

In this case, one can consider the **mean function***:*

$$\begin{array}{rccl} m: & S & \rightarrow & \mathbb{R} \\ & s & \mapsto & m(s) = \mathbb{E}X_s \end{array}$$

and the **covariance function***:*

$$\begin{array}{rccl} c: & S \times S & \rightarrow & \mathbb{R} \\ & (s,t) & \mapsto & c(s,t) = Cov(X_s, X_t). \end{array}$$

Proposition

A covariance function is **positive semidefinite (p.s.d.)**, *i.e.*

$$\forall n \geq 1, \forall (s_1, \ldots, s_n) \in S^n \text{ and } \forall a = (a_1, \ldots, a_n) \in \mathbb{R}^n,$$

we have:

$$\sum_{i=1}^{n} \sum_{j=1}^{n} a_i a_j c(s_i, s_j) \geq 0.$$

Proposition

The covariance function is said **positive definite (p.d.)** *if*

$$\forall n \geq 1 \text{ and } \forall (s_1, \ldots, s_n) \in S^n,$$

we have:

$$\sum_{i=1}^{n} \sum_{j=1}^{n} a_i a_j c(s_i, s_j) = 0 \iff (a_i = 0, \forall i = 1, \ldots, n).$$

Definition

A **gaussian random field** *X on S is a process such that, for all finite subset Λ of S and all sequence of reals $a = (a_s, s \in \Lambda)$, the r.v. $\sum_{s \in \Lambda} a_s X_s$ has a gaussian distribution.*

Definition

A second-order random field X on S is said to be **stationary** *if it has a constant mean function and its covariance function is invariant by translation, i.e.:*

$$\forall s \in S : m(s) = m$$
$$\forall (s,t) \in S^2, \forall h \in S : c(s+h, t+h) = c(s,t).$$

If X is stationary, we have: $c(s,t) = c(0, t-s)$, for all $(s,t) \in S^2$.

Definition

If X is stationary, the function

$$\begin{array}{rccl} C: & S & \rightarrow & \mathbb{R} \\ & h & \mapsto & C(h) = c(0,h) \end{array}$$

is called the **stationary covariance** *function.*

Definition

The **stationary correlation** *function of a stationary random field X is:*

$$\begin{array}{rccl} \rho: & S & \rightarrow & \mathbb{R} \\ & h & \mapsto & \rho(h) = \frac{C(h)}{C(0)}. \end{array}$$

Proposition

Let C be the stationary function of second-order spatial process. Then:

1. $C(h) = C(-h)$ *(even function)*
2. $\forall h \in S : |C(h)| \leq C(0)$ *(bounded function)*
3. *If C is continuous at the origin, then it is uniformly continuous on S.*

Proposition

Let C be the stationary function of second-order spatial process. Then:

$$\forall n \geq 1, \forall a \in \mathbb{R}^n, \forall (s_1, \ldots, s_n) \in S^n :$$
$$\sum_{i=1}^{n} \sum_{j=1}^{n} a_i a_j C(s_i - s_j) \geq 0 \ (s.d.p.).$$

Reciprocally, all s.d.p. function C is the covariance function of a stationary spatial process.

Proposition

Let C be the stationary function of a second-order spatial process. Then:

1. *If A is a linear function from $\mathbb{R}^d$ to $\mathbb{R}^d$, the random field $X^A = \{X_{As}, s \in S\}$ is stationary with covariance function $C^A(s) = C(As)$.*
   Moreover, if C is d.p. and A with full rank, then C^A is also d.p.
2. *If $C_1, \ldots, C_n, \ldots$ are stationary functions, then*
   - $\forall(\alpha_1, \alpha_2) \in \mathbb{R}^+ \times \mathbb{R}^+$ *the function* $C(h) = \alpha_1 C_1(h) + \alpha_2 C_2(h)$ *is a stationary covariance function*
   - $C(h) = C_1(h)C_2(h)$ *is a stationary covariance function*
   - $\lim_{n\to+\infty} C_n(h)$ *is also a stationary covariance function.*

Definition

A spatial process X is said to be **strictly stationary** *if: $\forall k \in \mathbb{N}, \forall (t_1, \ldots, t_k) \in S^k$ and $\forall h \in S$, the distribution of the random vector $(X_{t_1+h}, \ldots, X_{t_k+h})$ is independent of h.*

Let $||x|| = \sqrt{\sum_{i=1}^d x_i^2}$ denotes the euclidean norm on $\mathbb{R}^d$.

Definition

A second-order spatial process X has an **isotropic covariance function** *if $Cov(X_s, X_t)$ depends only on $||t - s||$, i.e. if there exists à function C_0 from $\mathbb{R}^+$ to $\mathbb{R}$ such that*

$$c(s,t) = C_0(||s - t||),$$

for all $(s,t) \in S^2$.

Definition

A spatial process X is said to be **intrinsically stationary** *or* **intrinsic** *if the processes*

$$\Delta X^h = \{\Delta X_s^h = X_{s+h} - X_s; s \in S\}$$

are stationary, for all $h \in S$.

One can show that if X is an intrinsic spatial process, then:

- there exists $m \in S$,called the drift, such that $\mathbb{E}(\Delta X_s^h) = \mathbb{E}(X_{s+h} - X_s) = \langle m, h \rangle$;
- there exists a function on S such that $\mathrm{Var}(\Delta X_s^h) = \mathrm{Var}(X_{s+h} - X_s) = 2\gamma(h)$

Without loss of generality, one can suppose the drift m to be equal to 0. This is why, one can find the simplified definition of an intrinsic process.

Definition

A spatial process X is said to be intrinsic if we have

$$\forall (s,h) \in S^2 : \mathbb{E}(X_{s+h} - X_s) = 0$$
$$\forall s \in S : Var(X_{s+h} - X_s) = 2\gamma(h).$$

The function γ is called the **semi-variogram** *function of X.*

Definition

The semi-variogram γ *of a spatial process* X *is said to be* **isotropic** *if there exists a function* γ_0 *such that:*

$$\gamma(h) = \gamma_0(||h||),$$

for all $h \in S$.

Proposition

If X *is a second order stationary process with covariance function* C, *then* X *is intrinsic with semi-variogram*

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

Proposition

The semi-variogram function γ of an intrinsic process X is such that:

1. $\gamma(h) = \gamma(-h)$ *(even function) and* $\gamma(0) = 0$;
2. *If A is a linear map on R^d, then the function $h \mapsto \gamma(Ah)$ is also a semi-variogram function;*
3. *If γ is continuous at 0, then γ is continuous at every s where γ is locally bounded.*
4. *If γ is bounded in the neighborhood of 0, then there exists positive reals a and b such that, for all $x \in S$:* $\gamma(x) \leq a||x||^2 + b$.

Definition

An **Allowable Linear Combination (A.L.C.)** *of a process X is a linear combination $\sum_{i=1}^n \lambda_i X_{s_i}$ of its coordinates with finite variance, i.e. such that*

$$Var\left(\sum_{i=1}^n \lambda_i X_{s_i}\right) < +\infty.$$

Proposition

If X is an intrinsic process, the linear combination $\sum_{i=1}^n \lambda_i X_{s_i}$ is an A.L.C. if, and only if, $\sum_{i=1}^n \lambda_i = 0$.

Proposition

The semi-variogram γ of an intrinsic process X is **conditionally negative definite**, *i.e. for all $n \in \mathbb{N}^*$, for all $a \in \mathbb{R}^n$ such that $\sum_{i=1}^n a_i = 0$ and for all $(s_1, \ldots, s_n) \in S^n$, we have:*

$$\sum_{i=1}^{n} \sum_{j=1}^{n} a_i a_j \gamma(s_i - s_j) \leq 0.$$

Theorem

A function γ defined on $\mathbb{R}^d$ is a semi-variogram if, and only if, this function is conditionally negative definite.

Proposition

If X is an intrinsic process with bounded semi-variogram, i.e. such that

$$\lim_{||h|| \to +\infty} \gamma(h) = \gamma(+\infty) < +\infty,$$

then X is second order stationary and

$$\gamma(+\infty) = C(0) = Var(X_S).$$

Theorem

A continuous function γ defined on $\mathbb{R}^d$ such that $\gamma(0) = 0$ is a semi-variogram if, and only if, for all $a > 0$, the function $h \mapsto e^{-a\gamma(h)}$ is a covariance function, i.e. is s.d.p.

Terminology

- When the limit

$$\lim_{||h|| \to +\infty} \gamma(h) = \gamma(+\infty) < +\infty,$$

  exists, its value $\gamma(+\infty)$ is called the **sill**.
- The **range** (resp. **practical range**) is the distance where (resp. 95% of) the value of the sill is reached.
- A semi-variogram has a **nugget effect** component when

$$\lim_{||h|| \to 0} \gamma(h) = \tau > 0.$$

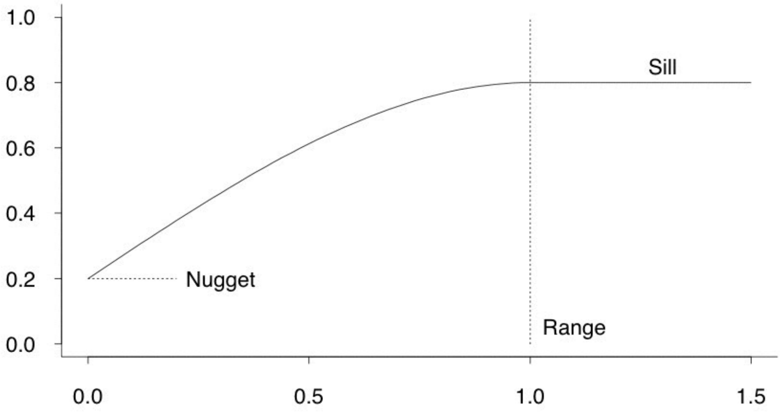

Figure: Nugget, Range and Sill of a Variogram

Examples of Isotropic variograms

C and a are always positive reals.

- **Pure nugget effect**
  - $$\gamma(h) = \begin{cases} 0 & \text{if} \quad h = 0 \\ C & \text{if} \quad h \neq 0 \end{cases}$$
  - Sill=Nugget effect= C
- **Exponential**
  - $$\gamma(h) = C \left(1 - \exp\left(-\frac{||h||}{a} \right) \right)$$
  - Sill= C
  - Practical Range= $3a$

- **Spherical** (when $d \leq 3$)
  - $$\gamma(h) = \begin{cases} C\left(\frac{3}{2}\frac{||h||}{a} - \frac{1}{2}\frac{||h||^3}{a^3}\right) & \text{if} \quad ||h|| \leq a \\ C & \text{if} \quad ||h|| > a \end{cases}.$$
  - Sill$=C$
  - Range$=a$
- **Gaussian**
  - $$\gamma(h) = C\left(1 - \exp\left(-\frac{||h||^2}{a^2}\right)\right)$$
  - Sill$=C$
  - Practical Range$=1.73a$
- **Generalized Exponential**

$$\gamma(h) = C\left(1 - \exp\left(-\frac{||h||^\alpha}{a^\alpha}\right)\right), \text{ for } \alpha \in]0,2]$$

- **Matern**

$$\gamma(h) = C\left(1 - \frac{1}{2^{\nu-1}\Gamma(\nu)}\left(\frac{||h||}{a}\right)^{\nu} K_\nu\left(\frac{||h||}{a}\right)\right), \text{ for } \nu > 1,$$

where K_ν is the modified Bessel function of the second kind.

- **Power**

$$\gamma(h) = C||h||^\alpha, \text{ for } \alpha \in]0, 2].$$

- **Mixed**, e.g.

$$\gamma(h) = \gamma_1(h) + \gamma_2(h) + \gamma_3(h),$$

where γ_1 is a pure nugget effect, γ_2 is spherical with low range and γ_3 is spherical with high range.

Anisotropy

Let $\overrightarrow{e}$ be a unit vector of $\mathbb{R}^d$: $||\overrightarrow{e}|| = 1$.

Definition

The **directional semi-variogram** $\gamma_{\overrightarrow{e}}$ *of a spatial process* X *in direction* $\overrightarrow{e}$ *is defined by:*

$$2\gamma_{\overrightarrow{e}}(h) = Var\left(X_{s+h\overrightarrow{e}} - X_s\right), \text{ for all } h \in \mathbb{R}.$$

Definition

A random field X *is said* **anisotropic** *if at least two of its directional semi-variogram differ.*

Definition

The semi-variogram γ of a random field X has a **geometric anisotropy** *if it results from a linear transformation A of an isotropic semi-variogram:*

$$\gamma(h) = \gamma_0(||Ah||) = \gamma(\sqrt{h^t Qh}), \text{ where } Q = A^t A.$$

Definition

The semi-variogram $h \mapsto \gamma(h)$ of a random field X has a **support anisotropy** *if its depends only on certain coordinates of h, possibly after a change of coordinates.*

Definition

The semi-variogram $h \mapsto \gamma(h)$ *of a random field* X *has a* **stratified (or zonal) anisotropy** *if it can be written as the sum of semi-variograms with different support anisotropies.*