Chapter 1
Second-order spatial models and geostatistics

Suppose $S \subseteq \mathbb{R}^d$ is a spatial set. A random field $X$ on $S$ taking values in a state space $E$ means a collection $X = \{X_s, s \in S\}$ of random variables (r.v.) indexed by $S$ taking values in $E$. This chapter is devoted to the study of second-order random fields, i.e., real-valued random fields where each $X_s$ has finite variance. We also study the broader class of intrinsic random fields, that is, random fields with increments of finite variance. We consider two approaches.

In the geostatistics approach, $S$ is a continuous subset of $\mathbb{R}^d$ and we model $X$ in a “second-order” way with its covariance function or its variogram. For example, for $d = 2$, $s = (x, y) \in S$ is characterized by fixed geographic coordinates and if $d = 3$, we add altitude (or depth) $z$. Spatio-temporal evolution in space can also be modeled at space-time “sites” $(s, t) \in \mathbb{R}^3 \times \mathbb{R}^+$, where $s$ represents space and $t$ time. Initially developed for predicting mineral reserves in an exploration zone $S \subseteq \mathbb{R}^3$, geostatistics is today used in a variety of domains (cf. Chilès and Delfiner (43); Diggle and Ribeiro (63)). These include, among others, earth science and mining exploration (134; 152), epidemiology, agronomy and design of numerical experiments (193). A central goal of geostatistics is to predict $X$ by kriging over all of $S$ using only a finite number of observations.

The second approach involves autoregressive (AR) models, used when $S$ is a discrete network of sites (we will also use the word “lattice”). $S$ may have a regular form, for example $S \subseteq \mathbb{Z}^d$ (images, satellite data, radiography; (42), (224)) or it may not (econometrics, epidemiology; (45), (7), (105)). Here, the spatial correlation structure is induced by the AR model chosen. Such models are well adapted to situations where measurements have been aggregated over spatial zones: for example, in econometrics this might be the percentages of categories of a certain variable in an administrative unit, in epidemiology, the number of cases of an illness per district $s$ and in agronomy, the total production in each parcel of land $s$. 
1.1 Some background in stochastic processes

Let \((\Omega, \mathcal{F}, \mathbb{P})\) be a probability space, \(S\) a set of sites and \((E, \mathcal{E})\) a measurable state space.

**Definition 1.1.** Stochastic process

A stochastic process (or process or random field) taking values in \(E\) is a family \(X = \{X_s, s \in S\}\) of random variables defined on \((\Omega, \mathcal{F}, \mathbb{P})\) and taking values in \((E, \mathcal{E})\). \((E, \mathcal{E})\) is called the state space of the process and \(S\) the (spatial) set of sites at which the process is defined.

For any integer \(n \geq 1\) and \(n\)-tuple \((s_1, s_2, \ldots, s_n) \in S^n\), the distribution of \((X_{s_1}, X_{s_2}, \ldots, X_{s_n})\) is the image of \(\mathbb{P}\) under the mapping \(\omega \mapsto (X_{s_1}(\omega), X_{s_2}(\omega), \ldots, X_{s_n}(\omega))\): that is, for \(A_i \in \mathcal{E}, i = 1, \ldots, n,\)

\[P_X(A_1, A_2, \ldots, A_n) = \mathbb{P}(X_{s_1} \in A_1, X_{s_2} \in A_2, \ldots, X_{s_n} \in A_n).\]

The event \((X_{s_1} \in A_1, X_{s_2} \in A_2, \ldots, X_{s_n} \in A_n)\) of \(\mathcal{E}\) is a cylinder associated with the \(n\)-tuple \((s_1, s_2, \ldots, s_n)\) and events \(A_i, i = 1, \ldots, n\) belonging to \(\mathcal{F}\). The family of all finite-dimensional distributions of \(X\) is called the spatial distribution of the process; if \(S \subseteq \mathbb{R}\), we say time distribution. More generally, the distribution of the process is uniquely defined as the extension of the spatial distribution to the sub-\(\sigma\)-algebra \(\mathcal{A} \subseteq \mathcal{F}\) generated by the set of cylinders of \(\mathcal{E}\) (32, Ch. 12), (180, Ch. 6).

For the rest of the chapter, we will be considering real-valued processes, \(E \subseteq \mathbb{R}\) endowed with a Borel \(\sigma\)-field \(\mathcal{E} = \mathcal{B}(E)\).

**Definition 1.2.** Second-order process

\(X\) is a second-order process (random field) if for all \(s \in S\), \(E(X_s^2) < \infty\). The mean of \(X\) (which necessarily exists) is the function \(m : S \rightarrow \mathbb{R}\) defined by \(m(s) = E(X_s)\).

The covariance of \(X\) is the function \(c : S \times S \rightarrow \mathbb{R}\) defined for all \(s, t\) by \(c(s, t) = \text{Cov}(X_s, X_t)\).

With \(L^2 = L^2(\Omega, \mathcal{F}, \mathbb{P})\) representing the set of real-valued and square integrable random variables on \((\Omega, \mathcal{F})\), \(X \in L^2\) means that \(X\) is a second-order process. A process \(X\) is said to be centered if for all \(s\), \(m(s) = 0\).

Covariances are characterized by the positive semidefinite (p.s.d.) property:

\[\forall m \geq 1, \forall a \in \mathbb{R}^m \text{ and } \forall (s_1, s_2, \ldots, s_m) \in S^m : \sum_{i=1}^{m} \sum_{j=1}^{m} a_i a_j c(s_i, s_j) \geq 0.\]

This property is a consequence of non-negativity of the variance of linear combinations:

\[\text{Var} \left( \sum_{i=1}^{m} a_i X_{s_i} \right) = \sum_{i=1}^{m} \sum_{j=1}^{m} a_i a_j c(s_i, s_j) \geq 0.\]

We say that the covariance is positive definite (p.d.) if furthermore, for every \(m\)-tuple of distinct sites, \(\sum_{i=1}^{m} \sum_{j=1}^{m} a_i a_j c(s_i, s_j) > 0\) whenever \(a \neq 0\). Gaussian processes are an important class of \(L^2\) processes.
1.2 Stationary processes

Definition 1.3. Gaussian process

\( X \) is a Gaussian process on \( S \) if for every finite subset \( \Lambda \subset S \) and real-valued sequence \( a = (a_s, s \in \Lambda) \), \( \sum_{s \in \Lambda} a_s X_s \) is a Gaussian random variable.

If \( m_A = E(X_A) \) is the mean of \( X_A = (X_s, s \in \Lambda) \) and \( \Sigma_A \) its covariance, then if \( \Sigma_A \) is invertible, the density (or likelihood) of \( X_A \) with respect to the Lebesgue measure on \( \mathbb{R}^{\sharp \Lambda} \) is

\[
 f_A(x_A) = (2\pi)^{-\sharp \Lambda/2} (\det \Sigma_A)^{-1/2} \exp \left\{ -\frac{1}{2} (x_A - m_A)^T \Sigma_A^{-1} (x_A - m_A) \right\},
\]

where \( \sharp U \) is the cardinality of \( U \) and \( x_A \) possible values of \( X_A \). Such densities are well-defined and Kolmogorov’s theorem ensures that for any mean function \( m \) and p.d. covariance \( c \) there exists a (Gaussian) random field with mean \( m \) and covariance \( c \).

Example 1.1. Brownian motion on \( \mathbb{R}^+ \) and Brownian sheet on \( (\mathbb{R}^+)^2 \)

\( X \) is a Brownian motion (180) on \( S = \mathbb{R}^+ \) if \( X_0 = 0 \), if for all \( s > 0 \), \( X_s \) follows a \( \mathcal{N}(0,s) \) \((X_s \sim \mathcal{N}(0,s))\) and if increments \( X([s,t]) = X_t - X_s, t > s \geq 0 \) are independent for disjoint intervals. The covariance of Brownian motion is \( c(s,t) = \min\{s,t\} \) and the increment process \( \Delta X_t = X_{t+\Delta} - X_t, t \geq 0 \) is stationary (cf. Ch. 1.2) with marginal distribution \( \mathcal{N}(0,\Delta) \).

This definition can be extended to the Brownian sheet (37) on the first quadrant \( S = (\mathbb{R}^+)^2 \) with: \( X_{u,v} = 0 \) if \( u \times v = 0 \), \( X_{u,v} \sim \mathcal{N}(0,u \times v) \) for all \( (u,v) \in S \) and independence of increments for disjoint rectangles; the increment on rectangle \([s,t], s = (s_1,s_2), t = (t_1,t_2), s_1 < t_1, s_2 < t_2 \) is given by

\[
 X([s,t]) = X_{t_1,t_2} - X_{t_1,s_2} - X_{s_1,t_2} + X_{s_1,s_2}.
\]

Brownian sheets are centered Gaussian processes with covariance \( c(s,t) = \min\{s_1,s_2\} \times \min\{t_1,t_2\} \).

1.2 Stationary processes

In this section, we suppose that \( X \) is a second-order random field on \( S = \mathbb{R}^d \) or \( \mathbb{Z}^d \) with mean \( m \) and covariance \( c \). The notion of stationarity of \( X \) can be more generally defined when \( S \) is an additive subgroup of \( \mathbb{R}^d \): for example, \( S \) could be the triangular lattice of \( \mathbb{R}^2 \), \( S = \{n e_1 + m e_2, n \text{ and } m \in \mathbb{Z}\} \) with \( e_1 = (1,0) \) and \( e_2 = (1/2, \sqrt{3}/2) \); another example is the finite \( d \)-dimensional torus with \( p^d \) points, \( S = (\mathbb{Z}/p\mathbb{Z})^d \).

1.2.1 Definitions and examples

Definition 1.4. Second-order stationary process

\( X \) is a second-order stationary process on \( S \) if it has constant mean and translation-invariant covariance \( c \):
\forall s,t \in S: E(X_s) = m \quad \text{and} \quad c(s,t) = \text{Cov}(X_s, X_t) = C(t - s).

\( C : S \to \mathbb{R} \) is the stationary covariance function of \( X \). Translation-invariance of \( c \) means:

\( \forall s,t,h \in S: c(s + h, t + h) = \text{Cov}(X_{t+h}, X_{t+h}) = C(s - t). \)

The correlation function of \( X \) is the function \( h \mapsto \rho(h) = C(h)/C(0) \). The following properties hold:

**Proposition 1.1.** Let \( X \) be a second-order stationary process with stationary covariance \( C \). Then:

1. \( \forall h \in S, \ |C(h)| \leq C(0) = \text{Var}(X_s). \)
2. \( \forall m \geq 1, a \in \mathbb{R}^m \) and \( \{t_1, t_2, \ldots, t_m\} \subseteq S: \sum_{i=1}^{m} \sum_{j=1}^{m} a_i a_j C(t_i - t_j) \geq 0. \)
3. If \( A : \mathbb{R}^d \longrightarrow \mathbb{R}^d \) is linear, the process \( X^A = \{X_{As}, s \in S\} \) is stationary with covariance \( C^A(s) = C(As) \). \( C^A \) is p.d. if \( C \) itself is and if \( A \) has full rank.
4. If \( C \) is continuous at the origin, then \( C \) is everywhere uniformly continuous.
5. If \( C_1, C_2, \ldots \) are stationary covariances, the following functions are as well:

   a. If \( C(h) = a_1 C_1(h) + a_2 C_2(h) \) if \( a_1 \) and \( a_2 \geq 0. \)

   b. More generally, if \( C(\cdot; u), u \in U \subseteq \mathbb{R}^k \) is a stationary covariance for each \( u \)

   and if \( \mu \) is a positive measure on \( \mathbb{R}^k \) such that \( C_\mu(h) = \int_U C(h; u) \mu(du) \) exists

   for all \( h \), then \( C_\mu \) is a stationary covariance.

   c. \( C(h) = C_1(h)C_2(h) \).

   d. \( C(h) = \lim_{n \to \infty} C_n(h) \), provided that the limit exists for all \( h \).

**Proof.** Without loss of generality, suppose that \( X \) is centered.

(1) is a consequence of the Cauchy-Schwarz inequality:

\[ C(h)^2 = \{E(X_h X_0)\}^2 \leq \{E(X_h^2)E(X_0^2)\} = E(X_h^2)^2. \]

(2) follows from the fact that covariances are p.s.d. (3) can be shown directly. (4) can be inferred from the fact that \( C(s + h) - C(s) = E[X_0(X_{s+h} - X_s)] \) and the Cauchy-Schwarz inequality,

\[ |C(s + h) - C(s)| \leq \sqrt{C(0)} \sqrt{2[C(0) - C(h)]}. \]

(5) It is easy to show that the functions \( C \) defined by (a), (b) and (d) are p.s.d. Then, if \( X_1 \) and \( X_2 \) are stationary and independent with covariances \( C_1 \) and \( C_2 \), covariance \( C \) given in (5-a) (resp. (5-b)) is that of \( X_i = \sqrt{a_1} X_{1i} + \sqrt{a_2} X_{2i} \) (resp. \( X_i = X_{1i}X_{2i} \)).

The notion of stationarity can defined in two ways in \( L^2 \). The first, weaker, is that of stationary increment processes or **intrinsic processes** and is presented in Section 1.3. The second, stronger, is known as **strict stationarity**. We say that \( X \) is strictly stationary if for all \( k \in \mathbb{N} \), all \( k \)-tuples \( (t_1, t_2, \ldots, t_k) \in S^k \) and all \( h \in S \), the distribution of \( (X_{t_1+h}, X_{t_2+h}, \ldots, X_{t_k+h}) \) is independent of \( h \). In a sense, \( X \) is strictly stationary if the spatial distribution of the process is translation-invariant.
If $X$ is strictly stationary and if $X \in L^2$, then $X$ is stationary in $L^2$. The converse is generally not true but both notions represent the same thing if $X$ is a Gaussian process.

**Example 1.2. Strong White Noise (SWN) and Weak White Noise (WWN)**

$X$ is a Strong White Noise if the variables $\{X_s, s \in S\}$ are centered, independent and identically distributed (i.i.d.). $X$ is a Weak White Noise if the variables $\{X_s, s \in S\}$ are centered and uncorrelated with finite constant variance: if $s \neq t$, $\text{Cov}(X_s, X_t) = 0$ and $\text{Var}(X_s) = \sigma^2 < \infty$. A SWN on $S$ is strictly stationary; a WWN on $S$ is a stationary process in $L^2$.

We denote $\|\cdot\|$ the Euclidean norm in $\mathbb{R}^d$: $\|x\| = \|x\|_2 = \sqrt{\sum_{i=1}^{d} x_i^2}$, $x = (x_1, x_2, \ldots, x_d)$.

**Definition 1.5. Isotropic covariance**

$X$ has isotropic covariance if for each $s, t \in S$, $\text{Cov}(X_s, X_t)$ depends only on $\|s - t\|$: 

$$\exists C_0 : \mathbb{R}^+ \rightarrow \mathbb{R} \text{ s.t.: } \forall t, s \in S, c(s, t) = C_0(\|s - t\|) = C(s - t).$$

Isotropic covariances are therefore stationary but isotropy imposes restrictions on the covariance. For example, if $X$ is isotropic and centered in $\mathbb{R}^d$ and if we consider $d + 1$ points mutually separated by distance $\|h\|$, 

$$E \left\{ \sum_{i=1}^{d+1} X_s \right\}^2 = (d + 1)C_0(\|h\|)(1 + d \rho_0(\|h\|)) \geq 0,$$

where $\rho_0 : \mathbb{R}^+ \rightarrow [-1, 1]$ is the isotropic correlation function. Therefore, for all $h$, this correlation satisfies 

$$\rho_0(\|h\|) \geq -1/d.$$ (1.1)

### 1.2.2 Spectral representation of covariances

Fourier theory and Bochner’s theorem (29; 43) together imply a bijection between stationary covariances $C$ on $S$ and their spectral measure $F$. It is thus equivalent to characterize a stationary model in $L^2$ by its stationary covariance $C$ or its spectral measure $F$.

**The $S = \mathbb{R}^d$ case**

We associate with $C$ a symmetric measure $F \geq 0$ bounded on the Borel sets $\mathcal{B}(\mathbb{R}^d)$ such that: 

$$C(h) = \int_{\mathbb{R}^d} e^{ihu} F(du),$$ (1.2)
where \( t^* h u = \sum_{i=1}^{d} h_i u_i \). If \( C \) is integrable, \( F \) is absolutely continuous with density \( f \) (with respect to the Lebesgue measure \( \nu \) on \( \mathbb{R}^d \)). \( f \) is called the spectral density of \( X \). The inverse Fourier transform lets us express \( f \) in terms of \( C \):

\[
f(u) = (2\pi)^{-d} \int_{\mathbb{R}^d} e^{-i t^* h u} C(h) dh.
\]

If \( X \) has isotropic covariance \( C \), its spectral density \( f \) does too and vice versa. Denote \( r = \|h\|, \ h = (r, \theta) \) where \( \theta = h \|h\|^{-1} \in S_d \) gives the direction of \( h \) in the unitary sphere \( S_d \) in \( \mathbb{R}^d \) centered at 0, \( \rho = \|u\| \) and \( u = (\rho, \alpha) \), with \( \alpha = u \|u\|^{-1} \in S_d \). For the polar coordinates \( h = (r, \theta) \) and \( u = (\rho, \alpha) \) of \( h \) and \( u \), note \( c_d(r) = C(h) \) and \( f_d(\rho) = f(u) \) the covariance and isotropic spectral density. Integrating (1.2) over \( S_d \) with surface measure \( d\sigma \), then over \( \rho \in [0, \infty[ \), we get:

\[
C(h) = c_d(r) = \int_{[0, \infty[} \left( \int_{S_d} \cos(r \rho \theta \alpha) d\sigma(\alpha) \right) \rho^{d-1} f_d(\rho) d\rho
\]

\[
= \int_{[0, \infty[} \Lambda_d(r \rho) \rho^{d-1} f_d(\rho) d\rho. \tag{1.3}
\]

The Hankel transform \( f_d \mapsto c_d \), analogous to a Fourier transform when dealing with isotropy shows that the variety of isotropic covariances is the same as that of the bounded positive measures on \([0, \infty[\). Furthermore (227), \( \Lambda_d(\nu) = \Gamma(d/2)(\nu/2)^{-d/2} \mathcal{I}_{(d-2)/2}(\nu) \), where \( \mathcal{I}_\kappa \) is the Bessel function of the first kind of order \( \kappa \) (2). For \( n = 1, 2 \) and 3, we have:

\[
\begin{align*}
 c_1(r) &= 2 \int_{[0, \infty[} \cos(\rho r) f_1(\rho) d\rho, \\
 c_2(r) &= 2\pi \int_{[0, \infty[} \rho J_0(\rho r) f_2(\rho) d\rho, \\
 c_3(r) &= \frac{2}{r} \int_{[0, \infty[} \rho \sin(\rho r) f_3(\rho) d\rho.
\end{align*}
\]

Using (1.3), we obtain lower bounds:

\[
C(h) \geq \inf_{\nu \geq 0} \Lambda_d(\nu) \int_{[0, \infty[} \rho^{d-1} f_d(\rho) d\rho = \inf_{\nu \geq 0} \Lambda_d(\nu) C(0).
\]

In particular, we get the lower bounds (227; 184), tighter than those in (1.1):

\[
\begin{align*}
\rho_0(\|h\|) &\geq -0.403 \text{ in } \mathbb{R}^2, \\
\rho_0(\|h\|) &\geq -0.218 \text{ in } \mathbb{R}^3, \\
\rho_0(\|h\|) &\geq -0.113 \text{ in } \mathbb{R}^4 \text{ and } \rho_0(\|h\|) \geq 0 \text{ in } \mathbb{R}^n.
\end{align*}
\]

**Example 1.3.** Exponential covariances in \( \mathbb{R}^d \)

For \( t \in \mathbb{R}, \ \alpha, b > 0, C_0(t) = b \exp(-\alpha |t|) \) has the Fourier transform:
\[ f(u) = \frac{1}{2\pi} \int_{-\infty,\infty} be^{-\alpha|\xi|} e^{-iut} d\xi = \frac{\alpha b}{\pi(\alpha^2 + u^2)}. \]

As \( f \geq 0 \) is integrable over \( \mathbb{R} \), it is a spectral density and \( C_0 \) therefore a covariance on \( \mathbb{R} \). Also, as

\[ \int_{0,\infty} e^{-\alpha x} \mathcal{K}(ux)x^{\kappa+1} dx = \frac{2\alpha(2u)^\kappa \Gamma(\kappa + 3/2)}{\pi^{1/2}(\alpha^2 + u^2)^{\kappa + 3/2}}, \]

we see that

\[ \phi(u) = \frac{\alpha b \Gamma([d+1]/2)}{[\pi(\alpha^2 + u^2)]^{(d+1)/2}} \]

is an isotropic spectral density of a process on \( \mathbb{R}^d \) with covariance

\[ C(h) = C_0(\|h\|) = b \exp(-\alpha \|h\|). \]

For any dimension \( d \), \( C \) is therefore a covariance function, given the name exponential, with parameter \( b \) for the variance of \( X \) and \( a = \alpha^{-1} \) the range.

**The \( S = \mathbb{Z}^d \) case**

Note \( T^d = [0,2\pi]^d \) the \( d \)-dimensional torus. According to Bochner’s theorem, any stationary covariance \( C \) on \( \mathbb{Z}^d \) is associated with a measure \( F \geq 0 \) bounded on the Borel sets \( \mathcal{B}(T^d) \) such that:

\[ C(h) = \int_{T^d} e^{iuh} F(du). \]

If \( C \) is square summable (\( \sum_{h \in \mathbb{Z}^d} |C(h)|^2 < \infty \)), the spectral measure \( F \) is absolutely continuous with density \( f \) (w.r.t. the Lebesgue measure) in \( L^2(T^d) \):

\[ f(u) = (2\pi)^{-d} \sum_{h \in \mathbb{Z}^d} C(h) e^{-iuh}. \] (1.4)

Furthermore, if \( \sum_{h \in \mathbb{Z}^d} |C(h)| < \infty \), we have uniform convergence and \( f \) is continuous. Also, the greater the differentiability of \( f \), the faster the convergence of \( C \) to 0 in the limit and vice versa: for example, if \( f \in C^k(T^d) \) where \( k = (k_1, \ldots, k_d) \in \mathbb{N}^d \),

\[ \lim_{h \to \infty} h^k |C(h)| < \infty, \]

where \( h = (h_1, h_2, \ldots, h_d) \to \infty \) means at least one coordinate \( h_i \to \infty \) and \( h^k = h_1^{k_1} \times \cdots \times h_d^{k_d} \). In particular, if \( f \) is infinitely differentiable, \( C \) goes to zero faster than any power function. This is the case for ARMA models (cf. §1.7.1) which have rational spectral density \( f \).
1.3 Intrinsic processes and variograms

1.3.1 Definitions, examples and properties

The stationarity property in $L^2$ may not be satisfied for various reasons: for example when $X_s = Y_s + Z$, where $Y$ is stationary in $L^2$ but $Z \notin L^2$, or equally when $X$ is in $L^2$ but not stationary, whether that be second-order (Brownian motion) or first-order ($X_s = a + b s + \varepsilon_s$ for a stationary centered residual process $\varepsilon$). A way to weaken the $L^2$ stationarity hypothesis is to consider the increment process $\{\Delta X_s(h) = X_{s+h} - X_s, s \in S\}$ of $X$, which may be stationary in $L^2$ even when $X$ is not stationary or not in $L^2$.

**Definition 1.6.** Intrinsic process

$X$ is an intrinsically stationary process (or intrinsic process) if for each $h \in S$, the process $\Delta X_s(h) = \{\Delta X_s(h) = X_{s+h} - X_s : s \in S\}$ is second-order stationary. The semi-variogram of $X$ is the function $\gamma : S \to \mathbb{R}$ defined by:

$$2\gamma(h) = \text{Var}(X_{s+h} - X_s).$$

Every stationary process in $L^2$ with covariance $C$ is clearly an intrinsic process with variogram $2\gamma(h) = 2(C(0) - C(h))$. However, the converse is not true: Brownian motion in $\mathbb{R}$, with variogram $|h|$, is intrinsic but not stationary. Furthermore, processes with affine means and stationary residuals are intrinsic, differentiation having the effect (as for time series) of absorbing affine trends and rendering the process first-order stationary. If we differentiate $k$ times, polynomial trends of degree $k$ can be removed, the process $X$ being called $k$-intrinsic if $\Delta^k X_s(h)$ is stationary (cf. (43); in $\mathbb{Z}$, so-called ARIMA models are a generalization of ARMA). For instance, the Brownian sheet on $(\mathbb{R}^+)^2$ is not intrinsic as it can be easily verified that $\text{Var}(X_{(u,v)} + (1,1) - X_{(u,v)}) = u + v + 1$ depends on $h = (u,v)$.

If $X$ is an intrinsic process and if the function $m(h) = E(X_{s+h} - X_s)$ is continuous at 0, then $m(\cdot)$ is linear: $\exists a \in \mathbb{R}^d$ s.t. $m(h) = \langle a, h \rangle$. In effect, $m$ is additive, $m(h) + m(h') = E\{(X_{s+h+h'} - X_{s+h'}) + (X_{s+h'} - X_s)\} = m(h + h')$ and continuity of $m$ at 0 implies linearity.

From now on, we will concentrate on intrinsic processes with centered increments: $\forall h, m(h) = 0$.

**Proposition 1.2.** Properties of variograms

1. $\gamma(h) = \gamma(-h), \gamma(h) \geq 0$ and $\gamma(0) = 0$.
2. Variograms are conditionally negative definite (c.n.d.): $\forall a \in \mathbb{R}^n$ s.t. $\sum_{i=1}^n a_i = 0, \forall \{s_1, \ldots, s_n\} \subseteq S$, we have:

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

3. If $A$ is a linear transformation in $\mathbb{R}^d$ and $\gamma$ a variogram, then $h \mapsto \gamma(Ah)$ is too.
4. Properties 5-(a,b,d) of covariances (cf. Prop. 1.1) remain true for variograms.
5. If \( \gamma \) is continuous at 0, then \( \gamma \) is continuous at every site \( s \) where \( \gamma \) is locally bounded.
6. If \( \gamma \) is bounded in a neighborhood of 0, \( \exists a \) and \( b \geq 0 \) such that for any \( x \), \( \gamma(x) \leq a\|x\|^2 + b. \)

**Proof.** (1) is obvious. To prove (2), set \( Y_s = (X_s - X_0) \). \( Y \) is stationary in \( L^2 \) with covariance \( C_Y(s, t) = \gamma(s) + \gamma(t) - \gamma(s-t) \). Then, if \( \sum_{i=1}^n a_i = 0 \), we get \( \sum_{i=1}^n a_iX_{s_i} = \sum_{i=1}^n a_iY_{s_i} \) and

\[
\text{Var} \left( \sum_{i=1}^n a_iX_{s_i} \right) = \sum_{i=1}^n \sum_{j=1}^n a_ia_jC_Y(s_i, s_j) = -\sum_{i=1}^n \sum_{j=1}^n a_ia_j\gamma(s_i - s_j) \geq 0.
\]

(3) If \( X \) is an intrinsic process with variogram \( 2\gamma \), then \( Y = \{Y_s = X_{As}\} \) is intrinsic with variogram:

\[
2\gamma(h) = \text{Var}(X_{A(s+h)} - X_{As}) = 2\gamma(Ah).
\]

(4) The proof is similar to that of Prop. 1.1. (5) \( 2\{\gamma(s + h) - \gamma(s)\} = E(A) \) where \( A = (X_{s+h} - X_0)^2 - (X_s - X_0)^2 \). It is easy to show that \( A = B + C \) where \( B = (X_{s+h} - X_s)(X_{s+h} - X_0) \) and \( C = (X_{s+h} - X_s)(X_s - X_0) \). Applying the Cauchy-Schwarz inequality to the products \( B \) and \( C \), the result follows from the upper bound:

\[
|\gamma(s + h) - \gamma(s)| \leq \sqrt{\gamma(h)}[\sqrt{\gamma(s)} + \sqrt{\gamma(s + h)}].
\]

Also, \( \gamma \) is uniformly continuous on any set over which \( \gamma \) is bounded. (6) We prove by induction that for each \( n \in \mathbb{N} \) and \( h = n \tilde{h} + \tau \), \( \gamma(nh) \leq n^2\gamma(h) \). This is true for \( n = 1 \); then, since

\[
2\gamma((n+1)h) = E\{(X_{s+(n+1)h} - X_{s+h}) + (X_{s+h} - X_s)\}^2,
\]

the Cauchy-Schwarz inequality gives

\[
\gamma((n+1)h) \leq \gamma(nh) + \gamma(h) + 2\sqrt{\gamma(nh)\gamma(h)} \leq \gamma(h)\{n^2 + 1 + 2n\} = (n + 1)^2\gamma(h).
\]

Suppose next that \( \delta > 0 \) satisfies \( \sup_{\|u\| \leq \delta} \gamma(u) = C < \infty \) and \( x \in \mathbb{R}^d \) satisfies \( n\delta \leq \|x\| \leq (n+1)\delta, \ n \geq 1 \). Setting \( \tilde{x} = \delta \|x\|^{-1} \), the decomposition \( x = n\tilde{x} + \tau \) defines some \( \tau \) satisfying \( \|\tau\| \leq \delta \). We conclude by remarking that

\[
\gamma(x) = \gamma(n\tilde{x} + \tau) \leq \gamma(n\tilde{x}) + \gamma(\tau) + 2\sqrt{\gamma(n\tilde{x})\gamma(\tau)} \leq Cn^2 + C + 2Cn = C(n + 1)^2 \leq C \left( \frac{\|x\|}{\delta} + 1 \right)^2.
\]

Unlike covariances, variograms are not necessarily bounded (for example, the variogram \( \gamma(h) = |h| \) for Brownian motion). However, the previous proposition shows that variograms tend to infinity at a rate of at most \( \|h\|^2 \). One such example
of quadratic growth $\gamma(t) = \sigma_1^2 t^2$ is that of the variogram of $X_t = Z_0 + tZ_1$, $t \in \mathbb{R}$, where $Z_0$ and $Z_1$ are centered and independent and $\text{Var}(Z_1) = \sigma_1^2 > 0$.

Characterizations exist to ensure a function $\gamma$ is a variogram, one of them being the following (43): if $\gamma$ is continuous and if $\gamma(0) = 0$, then $\gamma$ is a variogram if and only if, for every $u > 0$, $t \mapsto \exp\{-u\gamma(t)\}$ is a covariance. For example, as $t \mapsto \exp\{-u \|t\|^2\}$ is a covariance on $\mathbb{R}^d$ for each $u > 0$ and dimension $d$, $\gamma(t) = \|t\|^2$ is a variogram on $\mathbb{R}^d$ that goes to infinity at a quadratic rate.

### 1.3.2 Variograms for stationary processes

If $X$ is stationary with covariance $C$, then $X$ is intrinsic with variogram

$$2\gamma(h) = 2(C(0) - C(h)).$$

(1.5)

In particular, variograms of stationary processes are bounded. Matheron (153) partially proved the converse, that is, if the variogram of intrinsic process $X$ is bounded, then $X_t = Z_t + Y$ where $Z$ is a stationary process of $L^2$ and $Y$ some general real random variable.

If $C(h) \to 0$ as $\|h\| \to \infty$, then $\gamma(h) \to C(0)$ as $\|h\| \to \infty$. The variogram therefore has a sill at height $C(0) = \text{Var}(X)$ as $\|h\| \to \infty$. The range (resp. the practical range) is the distance at which the variogram reaches its sill (resp. 95% the value of the sill), cf. Fig. 1.1.

![Fig. 1.1](image_url) (a) Semivariogram of a stationary model with a nugget effect component; (b) variogram models that have the same range.

Statistical methods for second-order stationary processes can be considered in terms of covariances or in terms of variograms. Statisticians prefer the first way,
geostatisticians the second. We note that the advantage of working with variograms is that, unlike covariances, the mean does not have to be pre-estimated (cf. §5.1.4).

### 1.3.3 Examples of covariances and variograms

**Isotropic variograms**

The following examples are isotropic variograms on $\mathbb{R}^d$ traditionally used in geostatistics. Other models are presented in Yaglom (227), Chilès and Delfiner (43), Wackernagel (221) and the review article (195). The first five variograms, associated with stationary covariances $C(h) = C(0) - \gamma(h)$ are bounded with range parameter $a > 0$ and sill $\sigma^2$. Remember that $\| \cdot \|$ is the Euclidean norm on $\mathbb{R}^d$.

- **Nugget effect**: $\gamma(h; \sigma^2) = \sigma^2$ when $h > 0$, $\gamma(0) = 0$, associated with WWNs.

- **Exponential**: $\gamma(h; a, \sigma^2) = \sigma^2 \{1 - \exp(-\|h\|/a)\}$.

- **Spherical ($d \leq 3$)**:
  
  $$\gamma(h; a, \sigma^2) = \begin{cases} 
  \frac{\sigma^2}{2} \left\{1.5\|h\|/a - 0.5(\|h\|/a)^3\right\} & \text{if } \|h\| \leq a \\
  \sigma^2 & \text{if } \|h\| > a 
  \end{cases}$$

- **Generalized exponential, Gaussian**: $\gamma(h; a, \sigma^2, \alpha) = \sigma^2 (1 - \exp(-\|h\|/a)^\alpha)$ if $0 < \alpha \leq 2$; $\alpha = 2$ represents the Gaussian model.

- **Matérn**:
  
  $$\gamma(h; a, \sigma^2, \nu) = \sigma^2 \left\{1 - \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\|h\|/a\right)^\nu \mathcal{K}_\nu(\|h\|/a)\right\},$$

  where $\mathcal{K}_\nu(\cdot)$ is the modified Bessel function of the second kind with parameter $\nu > -1$ (2; 227; 200).

- **Power**: $\gamma(h; b, c) = b\|h\|^c$, $0 < c \leq 2$.

The variogram shown in Figure 1.1-(a) can be interpreted as being from a process $Y_s = X_s + \varepsilon_s$ where $\varepsilon$ is a white noise in $L^2$ (nugget effect at the origin) uncorrelated with $X$ whose variogram is continuous and with sill

$$2\gamma_Y(h) = 2\sigma^2_{\varepsilon}(1 - \delta_0(h)) + 2\gamma_X(h).$$

**Comments**

1. Spherical covariance can be interpreted in the following way: the volume $V(a, r)$ of the intersection of two spheres in $\mathbb{R}^3$ having the same diameter $a$ and centers at a distance $r$ apart is:
Fig. 1.2  Graph showing triangular, spherical and circular covariances with $\sigma^2 = 1$ and $a = 0.8$.

\[ V(a,r) = \begin{cases} 
\nu(S_a) \left\{ 1 - 1.5 \left( \frac{r}{a} \right) + 0.5 \left( \frac{r}{a} \right)^3 \right\} & \text{if } r \leq a \\
0 & \text{if } r > a 
\end{cases} \]

where $\nu(S_a)$ is the volume of a sphere of radius $a$. An example of a process leading to a spherical covariance is the process $X_s = N(S_a(s))$ counting the number of points of a homogeneous Poisson point process with intensity $\sigma^2 / \nu(S_a)$ in the sphere $S_a(s)$ of diameter $a$ centered at $s \in \mathbb{R}^3$ (cf. Ch. 3, §3.2).

2. The circular covariance $C_{circ}$ on $\mathbb{R}^2$ is obtained in the same way by replacing spheres in $\mathbb{R}^3$ by disks in $\mathbb{R}^2$:

\[ C_{circ}(h; a, \sigma^2) = \begin{cases} 
\frac{2\sigma^2}{\pi} \left( \arccos \left( \frac{\|h\|}{a} - \frac{\|h\|}{a} \sqrt{1 - \left( \frac{||h||}{a} \right)^2} \right) \right) & \text{if } ||h|| \leq a \\
0 & \text{otherwise} 
\end{cases} \]  \hspace{1cm} (1.6)

Similarly, the triangular covariance $C_{tri}$ on $\mathbb{R}^1$ can be obtained by simply replacing spheres in $\mathbb{R}^3$ by intervals $[-a, +a]$ in $\mathbb{R}^1$:

\[ C_{tri}(h; a, \sigma^2) = \begin{cases} 
\sigma^2 \left( 1 - \frac{|h|}{a} \right) & \text{if } |h| \leq a \\
0 & \text{otherwise} 
\end{cases} \]

Triangular, spherical and circular covariances are shown in Fig. 1.2.

3. As covariances on $\mathbb{R}^d$ remain positive semidefinite on any vectorial subspace, the restriction of a covariance to any subspace is still a covariance. In particular, the restriction of a spherical covariance to $\mathbb{R}^{d'}$, $d' \leq 3$, is still a covariance. However, extending an isotropic covariance from $\mathbb{R}^d$ to $\mathbb{R}^{d'}$ for $d' > d$ does not generally give a covariance. Exercise 1.5 gives an example of this with respect to the triangular covariance.
4. Our interest in Matérn covariance is due to its parameter $\nu$ which controls the variogram’s regularity at 0 (cf. Fig. 1.3), which in turn controls the quadratic mean (q.m.) regularity of the process $X$ (cf. §1.4) and its prediction $\hat{X}$ using kriging (cf. §1.9): increasing $\nu$ increases regularity of $\gamma$ at 0 and regularity of the process $X$ (the kriging surface $\hat{X}$). Taking $\nu = 1/2$ gives an exponential variogram which is continuous but not differentiable at 0, the associated process $X$ being continuous but not differentiable in q.m.; $\nu = \infty$ corresponds to the infinitely differentiable Gaussian variogram associated with an infinitely differentiable process $X$. For integer $m \geq 1$ and taking $\nu > m$, the covariance is differentiable $2m$ times at 0 and $X$ is differentiable $m$ times in q.m. For example, if $\nu = 3/2$ and $r = \|h\|$, $C(h) = C(r) = \sigma^2 (1 + (r/a)) \exp(-r/a)$ is twice differentiable at $r = 0$ and the associated random field differentiable in q.m.

5. The power model is self-similar, i.e., scale invariant: $\forall s > 0, \gamma(sh) = s^\alpha \gamma(h)$. It is therefore naturally associated with scale-free spatial phenomena and is the only model among those presented that has this property.

6. The generalized exponential model is identical to the exponential model when $\alpha = 1$ and the Gaussian model when $\alpha = 2$. Regularity of this type of variogram increases with $\alpha$ but the associated random field is only differentiable in quadratic mean when $\alpha = 2$.

7. Each of the previous models can be extended by taking positive linear combinations (or by integrating with respect to positive measures), in particular by adding a nugget effect variogram to any other variogram.

If $X$ is a sum of $K$ uncorrelated intrinsic processes (resp. stationary processes in $L^2$), it has the nested variogram (resp. covariance):
$2\gamma(h) = \sum_{j=1}^{K} 2\gamma_j(h)$ (resp. $C(h) = \sum_{j=1}^{K} C_j(h)$).

This model can be interpreted as having independent spatial components acting on different scales with different sills. Statistically speaking, small-scale components can only be identified if the sampling grid is fairly dense and large-scale components only if the diameter of the sampling domain in $S$ is relatively large.

### 1.3.4 Anisotropy

For a direction $\vec{e}$ in $\mathbb{R}^d$ such that $\|\vec{e}\| = 1$, the directional variogram of an intrinsic random field in direction $\vec{e}$ is defined as

$$2\gamma(h) = \text{Var}(X_{s+h\vec{e}} - X_s) \quad \text{for } h \in \mathbb{R}.$$ 

We say that a variogram is **anisotropic** if at least two directional varigrams differ.

We distinguish essentially two types of anisotropy: the first, geometric anisotropy is associated with a linear deformation of an isotropic model; the second corresponds to a nested variogram model over many subspaces of $\mathbb{R}^d$ (43; 77; 194).

**Geometric anisotropy**

The variogram $2\gamma$ on $\mathbb{R}^d$ exhibits geometric anisotropy if it results from an $A$-linear deformation of an isotropic variogram $2\gamma_0$:

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

i.e., if $\gamma(h) = \gamma_0(\sqrt{t}hQh)$, where $Q = 'AA$. Such variograms have the same sill in all directions (cf. Fig. 1.4-a) but with ranges that vary depending on the direction. In the orthonormal basis of eigenvectors of $Q$ associated with eigenvalues $(\lambda_k, k = 1, \ldots, d)$, $\gamma(\vec{h}) = \gamma_0(\sum_{k=1}^{d} \lambda_k \vec{h}_k)$ in these new coordinates $\vec{h}$.

For example, if $A$ is a rotation of angle $\phi$ around the origin in $\mathbb{R}^2$ followed by dilation by factor $0 \leq e \leq 1$ with respect to the new $y$ axis, the set of ranges forms an ellipse with eccentricity $e$ in this new basis. Figure 1.4-a gives an example of geometric anisotropy in $\mathbb{R}^2$ when $\gamma_0$ is an exponential model with parameters $a = 0.5$ and $\sigma^2 = 1$, with deformation $A$ the parameters $\phi = 45^0$ and $e = 0.7$.

We note that Sampson and Guttorp (192) propose a non-stationary model

$$\text{Var}(X_s - X_{s'}) = 2\gamma_0(g(s) - g(s')),$$

where $g$ is a bijective (or anamorphic) deformation of the space $S$ (cf. (170; 171) for examples of such deformations).
1.4 Geometric properties: continuity, differentiability

Stratified anisotropy

We talk of support anisotropy if variogram \( h \to 2\gamma(h) \), possibly after a change of coordinates, depends only on certain coordinates of \( h \): for example, if \( \mathbb{R}^d = E_1 \oplus E_2 \), where \( \dim(E_1) = d_1 \) and if \( 2\gamma_0 \) is an isotropic variogram on \( \mathbb{R}^{d_1} \), \( \gamma(h) = \gamma_0(h_1) \) for \( h = h_1 + h_2, h_1 \in E_1, h_2 \in E_2 \). The sill (and possibly the range) of \( \gamma \) will thus be direction-dependent (cf. Fig. 1.4-b). We say we have zonal anisotropy or stratified anisotropy if \( \gamma \) is the sum of several components, each with support anisotropy. For example,

\[
\gamma(h) = \gamma_1 \left( \sqrt{h_1^2 + h_2^2} \right) + \gamma_2(|h_2|)
\]

has a sill of height \( \sigma_1^2 + \sigma_2^2 \) in the \((0,1)\) direction and \( \sigma_i^2 \) in the \((1,0)\) direction, where \( \sigma_i^2 \) are the sills of \( \gamma_i, i = 1,2 \).

Chilès and Delfiner (43) suggest to avoid using separable models like \( \gamma(h) = \gamma_1(h_1) + \gamma_1(h_2) \) in \( \mathbb{R}^2 \) or \( \gamma(h) = \gamma_1(h_1,h_2) + \gamma_2(h_3) \) in \( \mathbb{R}^3 \) as certain linear combinations of \( X \) can end up with zero variance: for example, if \( X_s = X_s^1 + X_s^2 \), with \( \text{Cov}(X_s^1, X_s^2) = 0 \) and \( s = t(x,y) \), then \( \gamma(h) = \gamma_1(h_1) + \gamma_1(h_2) \) and for \( h_s = t(d_x,0) \), \( h_y = t(0,d_y), X_s - X_{s+h_s} - X_{s+h_y} + X_{s+h_s+h_y} \equiv 0 \).

More generally, anisotropy can be obtained by combining other anisotropies. Figure 1.4-b gives an example where \( \gamma_1 \) is the exponential model with geometric anisotropy and parameters \( a_1 = 0.5, \sigma_1^2 = 0.7, \phi = 45^0, e = 0.7 \) and \( \gamma_2 \) a different exponential model with parameters \( a_2 = 0.05, \sigma_2^2 = 0.3 \).

1.4 Geometric properties: continuity, differentiability

Let us now associate the set of \( L^2 \) processes with the following notion of mean square convergence:
Definition 1.7. Quadratic mean (q.m.) continuity

We say that a second-order process \( X = \{X_s, s \in S\} \) on \( S \subseteq \mathbb{R}^d \) is quadratic mean continuous at \( s \in S \) if for any converging sequence \( s_n \rightarrow s \) in \( S \), \( E(X_{s_n} - X_s)^2 \rightarrow 0 \).

The following proposition characterizes q.m. continuity.

Proposition 1.3. Let \( X \) be a centered \( L^2 \) process with covariance \( C(s,t) = \text{Cov}(X_s, X_t) \). Then \( X \) is everywhere q.m. continuous iff its covariance is continuous on the diagonal of \( S \times S \).

Proof. If \( C(s,t) \) is continuous at \( s = t = s_0 \), then \( E(X_{s_0+h} - X_{s_0})^2 \rightarrow 0 \) as \( h \rightarrow 0 \). In effect:

\[
E(X_{s_0+h} - X_{s_0})^2 = C(s_0+h, s_0) - 2C(s_0+h, s_0) + C(s_0, s_0).
\]

To show the converse, we write:

\[
\Delta = C(s_0 + h, s_0 + k) - C(s_0, s_0) = e_1 + e_2 + e_3,
\]

with \( e_1 = E[(X_{s_0+h} - X_{s_0})(X_{s_0+k} - X_{s_0})] \), \( e_2 = E[(X_{s_0+h} - X_{s_0})X_{s_0}] \) and \( e_3 = E[X_{s_0}(X_{s_0+k} - X_{s_0})] \). If \( X \) is q.m. continuous, then \( e_1, e_2 \) and \( e_3 \rightarrow 0 \) if \( h \) and \( k \rightarrow 0 \) and \( C \) is continuous on the diagonal. \( \square \)

Almost sure (a.s.) continuity of trajectories is a result of a different nature and much harder to obtain. We have for example the following result (3): if \( X \) is a centered Gaussian process with continuous covariance, a.s. continuity of trajectories on \( S \subseteq \mathbb{R}^d \) is assured if

\[
\exists c < \infty \text{ and } \epsilon > 0 \text{ s.t. } \forall s, t \in S: E(X_s - X_t)^2 \leq c |\log \|s - t\||^{-(1+\epsilon)}.
\]

When \( X \) is an intrinsic Gaussian process, this continuity holds if \( \gamma(h) \leq c |\log \|h\||^{-(1+\epsilon)} \) in a neighborhood of the origin. Apart from the nugget effect model, all variograms presented in §1.3.3 satisfy this property and the associated (Gaussian) models therefore have a.s. continuous trajectories.

We now examine differentiability in \( L^2 \) in given directions, or, equivalently, differentiability of processes in \( \mathbb{R}^1 \).

Definition 1.8. Quadratic mean differentiability

We say the process \( X \) on \( S \subseteq \mathbb{R}^1 \) is q.m. differentiable at \( s \) if there exists a real random variable (r.r.v.) \( \dot{X}_s \) such that

\[
\lim_{h \to 0} \frac{X_{s+h} - X_s}{h} = \dot{X}_s \text{ in } L^2.
\]

We note that all trajectories of a process \( X \) might be extremely regular without \( X \) being q.m. differentiable (cf. Ex. 1.11).

Proposition 1.4. Let \( X \) be a centered \( L^2 \) process with (not necessarily stationary) covariance \( C(s,t) = \text{Cov}(X_s, X_t) \). If \( \frac{\partial^2}{\partial s \partial t} C(s,t) \) exists and is finite on the diagonal
of \( S \times S \), then \( X \) is everywhere q.m. differentiable, the second-order mixed partial derivative \( \frac{\partial^2}{\partial s \partial t} C(s,t) \) exists everywhere and the covariance of the derived process is \( \text{Cov}(X_s, X_t) = \frac{\partial^2}{\partial s \partial t} C(s,t) \).

**Proof.** Let \( Y_s(h) = (X_{s+h} - X_s)/h \). To show existence of \( \dot{X}_s \), we use Loève’s criterion ((145), p. 135) which says that \( Z_h \to Z \) in \( L^2 \) iff \( E(Z_h Z_k) \to c < \infty \) whenever \( h \) and \( k \to 0 \) independently. Next let \( \Delta_{s,t}(h,k) = E(Y_s(h)Y_t(k)) \). It is easy to show that:

\[
\Delta_{s,t}(h,k) = h^{-1}k^{-1}\{C(s+h,t+k) - C(s,h,t) - C(s,t+k) + C(s,t)\}.
\] (1.7)

Therefore, if \( \frac{\partial^2}{\partial s \partial t} C(s,t) \) exists and is continuous at \((s,s)\),

\[
\lim_{h \to 0} \lim_{k \to 0} E(Y_s(h)Y_s(k)) = \frac{\partial^2}{\partial s \partial t} C(s,s).
\]

Loève’s criterion thus ensures convergence of \( (Y_s(h)) \) towards a limit denoted \( \dot{X}_s \) whenever \( h \to 0 \) and the process \( X = \{X_s, s \in S\} \) is in \( L^2 \). Let \( C^* \) be the covariance of \( \dot{X} \). Using (1.7), \( C^*(s,t) \) is the limit of \( \Delta_{s,t}(h,k) \) when \( h,k \to 0 \) and therefore \( \frac{\partial^2}{\partial s \partial t} C(s,t) = C^*(s,t) \) exists everywhere. \( \square \)

**1.4.1 Continuity and differentiability: the stationary case**

**Continuity**

We can deduce easily from the previous results that an intrinsic (resp. stationary in \( L^2 \)) process is q.m. continuous if its variogram \( 2\gamma \) (resp. covariance \( C \)) is continuous at \( h = 0 \); in such cases the variogram \( 2\gamma \) (resp. covariance \( C \)) is continuous on any set where \( \gamma \) is bounded (resp. everywhere continuous; cf. Prop. 1.2). Matérn (154) has shown more precisely that if a random field allows a variogram that is everywhere continuous except at the origin, then this random field is the sum of two uncorrelated random fields, one associated with a pure nugget effect and the other with an everywhere continuous variogram.

**Differentiability**

\( t \mapsto X_t \) is q.m. differentiable in \( \mathbb{R} \) if the second derivative \( \gamma''(0) \) of the variogram exists. In this case, the second derivative \( \gamma'' \) exists everywhere, \( X \) is stationary with covariance \( \gamma'' \) and the bivariate process \( (X,X) \in L^2 \) satisfies (227):
\[ E(\dot{X}_{s+\tau}X_s) = \gamma'(\tau) \quad \text{and} \quad E(\dot{X}_{s+\tau}\dot{X}_s) = -\gamma'(\tau). \]

In particular, as \( \gamma'(0) = 0 \), \( X_s \) and \( X_t \) are uncorrelated for all \( s \) and independent if \( X \) is a Gaussian process. We remark that if \( X \) is stationary and supposing \( C(s,t) = c(s-t), \quad C''(s,t) = -c''(s-t) \) and \( c''(0) \) exists, then \( X \) is stationary with covariance \(-c''\).

For integer \( m \geq 1 \), we say that \( X \) is \( m \)-th order q.m. differentiable if \( X^{(m-1)} \) exists in q.m. and if \( X^{(m-1)} \) is q.m. differentiable. If we suppose \( X \) is stationary with covariance \( C \), then \( X \) is \( m \)-th order differentiable if \( C^{(2m)}(0) \) exists and is finite. In this case, \( X^{(m)} \) is stationary with covariance \( t \mapsto (-1)^m C^{(2m)}(t) \). For example, a Matérn process is \( m \)-th order q.m. differentiable whenever \( v > m \) (200).

If \( \gamma \) is infinitely differentiable at the origin, \( X \) is infinitely q.m. differentiable. In this case, \( X_t = \lim_{t^2} \sum_{k=0}^n t^k X_0^{(k)} / k! \) (200). \( X \) is “purely deterministic” as it suffices to know \( X \) is a (small) neighborhood of 0 to know it everywhere. This may lead us to put aside an infinitely differentiable variogram model (i.e., the Gaussian variogram) if we are not sure about the deterministic nature and/or hyperregularity of \( X \).

**Example 1.4.** Quadratic mean regularity for processes on \( \mathbb{R}^2 \)

Figure 1.5 gives an idea of process regularity for three different variograms. Simulations were performed using the RandomFields package (cf. §4.7).

(a) \( X \) is a GWN (Gaussian WN) with a nugget effect variogram \( \gamma \) that is not continuous at 0. \( X \) is not q.m. continuous, trajectories are extremely irregular.

(b) \( \gamma \) is exponential, isotropic and linear at the origin: \( \gamma(h) = a + b ||h|| + o(||h||) \), continuous but not differentiable at 0. \( X \) is q.m. continuous but not differentiable.

(c) \( \gamma \) is a class \( \mathcal{C}^2 \) (in fact, \( \mathcal{C}^{\infty} \)) isotropic Gaussian variogram at the origin. Trajectories are q.m. continuous and differentiable. We would have the same regularity for any variogram of the form \( a + b \ ||h||^{\alpha} \) at the origin, for \( \alpha \geq 2 \).

**Example 1.5.** Separable covariances with cubic components

Separable covariances \( C(h) = \prod_{k=1}^d C_k(h_k) \), where \( h = (h_1, h_2, \ldots, h_d) \in \mathbb{R}^d \) are used for kriging due to ease of manipulation (cf. §1.9), particularly when performing simulations. They also help us to easily verify directional differentiability of the associated process. Separable covariances with cubic components (132) are associated with correlations in \([0,1]\) of the following type: for \( \rho, \gamma \) and \( h \in [0,1] \),

\[ C(h) = 1 - \frac{3(1-\rho)}{2+\gamma} h^2 + \frac{(1-\rho)(1-\gamma)}{2+\gamma} h^3. \]  

(1.8)

\( C \) is p.d. if \( \rho \geq (5\gamma^2 + 8\gamma - 1)(\gamma^2 + 4\gamma + 7)^{-1} \) (158). In this case, a process \( X \) with covariance \( C \) is q.m. differentiable and its derivative \( \dot{X} \) has affine covariance in \([0,1]\):

\[ C_{\dot{X}}(h) = -C''(h) = \tau^2 \{1 - (1-\gamma)h\}, \]
1.5 Spatial modeling using convolutions

1.5.1 Continuous model

A natural way to construct (Gaussian) models \( X = (X_s, s \in S) \) over subsets \( S \) of \( \mathbb{R}^d \) is to consider the convolution

\[
X_s = \int_{\mathbb{R}^d} k(u,s)W(du),
\]

where \( \tau^2 = 6(1 - \rho)/(2 + \gamma) \). Parameters \( \rho = \text{Cor}(X_0, X_1) = C(1) \) and \( \gamma = \text{Cor}(X_0, X_1) = C_{X, X} \) can be respectively interpreted as correlations between the final observations and between their derivatives.
where \( \mathcal{K} = \{ u \to k(u, s), s \in S \} \) is a family of non-random real-valued kernels on \( \mathbb{R}^d \) and \( W \) a centered latent (Gaussian) random field with orthogonal increments on \( \mathbb{R}^d \), that is: for \( \delta \) the Dirac delta function,

\[
E(W(du)W(dv)) = \delta(u, v)du \times dv.
\]

A classic choice for \( W \) when \( d = 1 \) is Brownian motion (in this case, the convolution is a Wiener integral) or the Brownian sheet when \( d = 2 \) (cf. Example 1.1, (37) and Ex. 1.14). Convolution (1.9) is well-defined in \( L^2 \) if for any \( s \in S \), \( k(\cdot, s) \) is square integrable (227, p. 67-69).

\( X_s \) is therefore a centered process with covariance:

\[
C(s, t) = \text{Cov}(X_s, X_t) = \int_S k(u, s)k(u, t)du.
\]

This model can be second-order characterized either by kernel family \( k \) or by its covariance \( C \) as \( X \) is a Gaussian process if \( W \) is. If \( S = \mathbb{R}^d \) and if kernel family \( k \) is translation-invariant, \( k(u, s) = k(u - s) \) and \( \int k^2(u)du < \infty \), then \( X \) is stationary with covariance

\[
C(h) = \text{Cov}(X_s, X_{s+h}) = \int_S k(u)k(u-h)du.
\]

If \( k \) is isotropic, \( X \) is too and the mapping between \( C \) and \( k \) is bijective. Examples of such mappings \( C \leftrightarrow k \) can be found in (219; 43, p. 646):

**Gaussian covariance \( C \), \( d \geq 1 \), \( a > 0 \):**

\[
k(u) = \sigma \exp\{-a\|u\|^2\} \leftrightarrow C(h) = \sigma^2 \left( \frac{\pi}{2a} \right)^{d/2} \exp\left\{-\frac{a}{2} \|h\|^2\right\};
\]

**Exponential covariance \( C \), \( d = 3 \), \( a > 0 \):**

\[
k(u) = 2\sigma a^{-1/2} \left( 1 - \frac{\|u\|}{a} \right) \exp\left( -\frac{\|h\|}{a} \right) \leftrightarrow C(h) = \sigma^2 \exp\left( -\frac{\|h\|}{a} \right);\]

**Spherical covariance \( C \), \( d = 3 \), \( a > 0 \):**

\[
k(u) = c1 \left\{ \|u\| \leq \frac{a}{2} \right\} \leftrightarrow C(h) = V_d \left( \frac{a}{2} \right) \left( 1 - \frac{3}{2} \frac{\|h\|}{a} + \frac{1}{2} \left( \frac{\|h\|}{a} \right)^3 \right) 1\{\|h\| < a\}.
\]

Such mappings are no longer bijective if \( X \) is stationary and non-isotropic as several different kernels \( k \) can give the same covariance \( C \).

We now describe several advantages of representing \( X \) using convolutions (112):

1. Formula (1.9) allows us to deal with all second-order models without having to satisfy the positive definiteness condition for covariances (219).
2. With (1.9) we can generate non-Gaussian models whenever the convolution is well-defined. For example, if \( W \) is a Poisson process (cf. §3.2) (resp. Gamma process (225)) with independent increments, convolution allows us to model a process \( X \) with values in \( \mathbb{N} \) (resp. values in \( \mathbb{R}^+ \)).
3. For non-stationary but slowly-varying kernel families $k$ we can propose parametrized types of non-stationarity for $X$ (cf. (113) for modeling environmental data).

4. If we choose the latent process $W$ to be a function of time $t$, convolution allows us to construct spatio-temporal models where the kernel $k$ may or may not be a function of time. For example, a time-dependent model with a kernel that is constant for given $t$ is $X_s(t) = \int_S k(u,s)W(du,t)$.

5. When observing a multivariate phenomenon $X \in \mathbb{R}^p$, multivariate convolution allows construction of spatially correlated components by choosing in (1.9) a kernel $k \in \mathbb{R}^p$. For example, if $S_0 \cup S_1 \cup S_2$ is a partition of $S$ (112),

$$X_{1,s} = \int_{S_0 \cup S_1} k_1(u-s)W(du) \quad \text{and} \quad X_{2,s} = \int_{S_0 \cup S_2} k_2(u-s)W(du).$$

### 1.5.2 Discrete convolution

In practice, we have to use discrete convolutions of $W$ at $m$ sites $\mathcal{U} = \{u_1, u_2, \ldots, u_m\}$ of $S$: $\mathcal{U}$ is a convolution support that allows us to get reasonably close to spatial integral (1.9). Denoting $w = ^t(w_1, w_2, \ldots, w_n)$ where $w_i = w(u_i)$, $i = 1, \ldots, m$, this model is written

$$X_s^w = (K * w)_s = \sum_{i=1}^{m} k(u_i, s)w_i, \quad s \in S,$$

(1.10)

where $w$ is a WWN with variance $\sigma_w^2$. Such models thus depend on the choice of support $\mathcal{U}$, though the spatial index $s$ remains continuous. Though this formulation can be interpreted as a moving average (MA, cf. §1.7.1), the difference here is that there is no notion of closeness between $s$ and the sites of $\mathcal{U}$, (1.10) being interpreted as an approximation to the continuous model (1.9).

Given $n$ observations $X = ^t(X_{s_1}, X_{s_2}, \ldots, X_{s_n})$ of $X$ at $O = \{s_1, s_2, \ldots, s_n\}$, a model that incorporates exogenous variables $z \in \mathbb{R}^p$ and a WWN measurement error $\varepsilon$ can be written, for each site,

$$X_s = ^t(z_s \beta + X_s^w + \varepsilon_s), \quad s \in O, \beta \in \mathbb{R}^p.$$

(1.11)

This can be put in matrix form as:

$$X = Z\beta + Kw + \varepsilon,$$

where $K = (K_{i,l})$, $K_{i,l} = k(u_i, s_l)$, $l = 1, \ldots, n$ and $i = 1, \ldots, m$. The model’s parameters are $\mathcal{U}, k(\cdot)$ and $(\beta, \sigma_w^2, \sigma_\varepsilon^2)$. Using statistics vocabulary, we say we are dealing with a random effects linear model where $w$ is the cause of the random effect $Kw$ and the deterministic trend is modeled using covariates $z$.

A possible choice for $\mathcal{U}$ is the regular triangular network with spacing $\delta$; $\delta$ is a compromise between giving a good data fit (small spacing) and simple calculations (larger spacing). A compromise is to use a multiresolution model with two or
more spacings. For example, the random component of a two-resolution model with triangular spacings $\delta$ and $\delta/2$ is written,

$$X^w = X^{1w} + X^{2w},$$

where $X^{1w}$ (resp. $X^{2w}$) is component (1.10) associated with this $\delta$-spacing and kernel $k_1$ (resp. $\delta/2$-spacing and kernel $k_2$).

In this context, a Bayesian formulation (cf. for example (143)) might be considered as it can incorporate uncertainty in the parameters determining the convolution.

Discrete convolutions equally allow us to construct non-stationary, non-Gaussian and multivariate models as well as spatio-temporal ones (208; 112). For example, (112) models the random component of the temporal evolution of ozone concentration over $T = 30$ consecutive days in a region of the United States by

$$X^w_s(t) = \sum k(u_i - s)w_i(t), \quad s \in \mathcal{S}, t = 1, \ldots, T,$$

with $\{w_i(t), t = 1, \ldots, T\}$ as $T$ independent Gaussian random walks on a spatial support $\mathcal{U}$ of 27 sites.

### 1.6 Spatio-temporal models

We now present several spatio-temporal geostatistics models. This subject is undergoing significant expansion, in particular for applications in climatology and environmental sciences (136; 133; 142; 91). (126; 36) give models derived from stochastic differential equations, (148; 223; 208; 112) develop discrete-time approaches and (202) compare discrete and continuous-time approaches.

Suppose $X = \{X_{s,t}, s \in \mathcal{S} \subseteq \mathbb{R}^d \text{ and } t \in \mathbb{R}_+\}$ is a real-valued process with $s$ representing space and $t$ time. $X$ is second-order stationary (resp. isotropic) if:

$$\text{Cov}(X_{s_1,t_1}, X_{s_2,t_2}) = C(s_1 - s_2, t_1 - t_2) \quad \text{(resp. } = C(\|s_1 - s_2\|, |t_1 - t_2|)).$$

As $(s,t) \in \mathbb{R}^d \times \mathbb{R} = \mathbb{R}^{d+1}$, one possible approach is to consider time as an additional dimension and to reuse definitions and model properties studied earlier, this time in dimension $d + 1$. However, this strategy does not take into account the fact that spatial and temporal variables work on different scales and have different meanings. For example, the isotropic exponential model $C(s,t) = \sigma^2 \exp\{-\|s,t\|/a\}$, where $s \in \mathbb{R}^d$ and $t \in \mathbb{R}$ is far from ideal; it is more natural to consider a geometric anisotropy model of the type $C(s,t) = \sigma^2 \exp\{-\|s\|/b + |t|/c\}$, $b,c > 0$. Proposition 1.1 then provides necessary tools to define more flexible stationary models in which spatial and temporal variables are treated separately. Furthermore, it may be pertinent to suggest semi-causal spatio-temporal models in which the concept of time has a well-defined meaning.
Covariances can be separable, as in these two examples:

(i) additive: \( C(s, t) = C_S(s) + C_T(t) \);
(ii) factorizing: \( C(s, t) = C_S(s)C_T(t) \),

where \( C_S(\cdot) \) is a spatial covariance and \( C_T(\cdot) \) a temporal one. Type (i) includes zonal anisotropy cases; these spatial and temporal anisotropies can be uncovered by performing variographic analysis (cf. §5.1.1) separately for space (considering pairs of sites \((s_1, s_2)\) at the same time \(t\)) and time (considering pairs of times \((t_1, t_2)\) at the same site \(s\)).

Separable space \times time covariance

Case (ii) covers what are known as covariances that are separable in space and time.

The advantage of separable models is that they simplify the calculation of the covariance matrix, its inverse and its spectrum when \( X \) is observed on the rectangle \( S \times T = \{s_1, s_2, \ldots, s_n\} \times \{t_1, t_2, \ldots, t_m\} \). More precisely, if \( X = (X_{s_1, t_1}, \ldots, X_{s_n, t_1}, \ldots, X_{s_1, t_m}, \ldots, X_{s_n, t_m}) \) is the vector of the \( n \times m \) observations, \( \Sigma = \text{Cov}(X) \) is the Kronecker product of \( \Sigma_T \), an \( m \times m \) temporal covariance matrix with \( \Sigma_S \), the \( n \times n \) spatial covariance matrix:

\[
\Sigma = \Sigma_T \otimes \Sigma_S.
\]

The product \( \Sigma \) is thus an \( mn \times mn \) matrix made up of \( m \times m \) blocks \( \Sigma_{k,l} \) of size \( n \times n \) equal to \( C_T(k - l) \Sigma_S \). The inverse and determinant of \( \Sigma \) are then easily calculated:

\[
(\Sigma)^{-1} = (\Sigma_T)^{-1} \otimes (\Sigma_S)^{-1}, \quad |\Sigma| = |\Sigma_T|n|\Sigma_S|m
\]

and the spectrum of \( \Sigma \) is the termwise product of the spectra of \( \Sigma_T \) and \( \Sigma_S \). These properties simplify prediction, simulation and estimation of such models, especially when the spatial \((n)\) or temporal \((m)\) domain of observation is large.

The downside of separable models is that they do not allow spatio-temporal interactions \( C_S(s_1 - s_2; u) \) between future instants of time \( u \) since \( C(s_1 - s_2, t_1 - t_2) = C_S(s_1 - s_2)C_T(u) \). Also, separability implies reflective symmetry \( C(s, t) = C(-s, -t) \) of the covariance, a condition that is not generally needed.

Non-separable models

Cressie and Huang (50) propose constructing a non-separable model using the spectral density \( g \):

\[
C(h, u) = \int_{\omega \in \mathbb{R}^d} \int_{\tau \in \mathbb{R}} e^{i(h\omega + u\tau)} g(\omega, \tau) d\omega d\tau. \quad (1.12)
\]

If we express \( g(\omega, \cdot) \) as the Fourier transform on \( \mathbb{R} \) of some function \( h(\omega, \cdot) \),

\[
g(\omega, \tau) = \frac{1}{2\pi} \int_{\mathbb{R}} e^{-iu\tau} h(\omega, u) du,
\]
where \( h(\omega, u) = \int_{\mathbb{R}} e^{i\omega \tau} g(\omega, \tau) d\tau \), the spatio-temporal covariance can be written

\[
C(h, u) = \int_{\mathbb{R}^d} e^{i h \omega} h(\omega, u) d\omega.
\]

However, we can always write:

\[
h(\omega, u) = k(\omega) \rho(\omega, u), \tag{1.13}
\]

where \( k(\cdot) \) is a spectral density on \( \mathbb{R}^d \) and where for each \( \omega \), \( \rho(\omega, \cdot) \) is an autocorrelation function on \( \mathbb{R} \). Thus, under the following conditions:

1. For each \( \omega \), \( \rho(\omega, \cdot) \) is a continuous autocorrelation function on \( \mathbb{R} \) satisfying
   \[
   \int_{\mathbb{R}} \rho(\omega, u) du < \infty \text{ and } k(\omega) > 0;
   \]
2. \( \int_{\mathbb{R}^d} k(\omega) d\omega < \infty \),

the function \( C \) defined as:

\[
C(h, u) = \int_{\mathbb{R}^d} e^{i h \omega} k(\omega) \rho(\omega, u) d\omega \tag{1.14}
\]

is a spatio-temporal covariance. If \( \rho(\omega, u) \) is independent of \( \omega \), this model is separable.

**Example 1.6.** The Cressie-Huang model: if we take,

\[
\rho(\omega, u) = \exp\left(-\frac{||\omega||^2 u^2}{4}\right) \exp\left(-\delta u^2\right), \quad \delta > 0 \text{ and}
\]

\[
k(\omega) = \exp\left(-\frac{c_0 ||\omega||^2}{4}\right), \quad c_0 > 0,
\]

then:

\[
C(h, u) \propto \frac{1}{(u^2 + c_0)^{d/2}} \exp\left(-\frac{||h||^2}{u^2 + c_0}\right) \exp\left(-\delta u^2\right). \tag{1.15}
\]

The condition \( \delta > 0 \) is needed to ensure that \( \int \rho(0, u) du < \infty \), but the limit as \( \delta \to 0 \) of (1.15) remains a spatio-temporal covariance function.

The weakness of this approach is that it requires calculation of Fourier transforms on \( \mathbb{R}^d \). Gneiting (90) proposes a different approach: let \( \psi(t), \; t \geq 0 \) be a strictly monotone function and \( t \mapsto \phi(t) > 0, \; t \geq 0 \) a function for which \( \phi'(t) \) is strictly monotone. Then the following function is a spatio-temporal covariance:

\[
C(h, u) = \frac{\sigma^2}{\phi(||u||^2) \psi\left(\frac{||h||^2}{\phi(||u||^2)}\right)}. \tag{1.16}
\]
Example 1.7. Gneiting’s spatio-temporal covariance

If \( \psi(t) = \exp(-ct^\gamma), \phi(t) = (at^{\alpha} + 1)^{\beta} \), with \( a \geq 0, c \geq 0, \alpha, \gamma \in [0, 1], \beta \in [0, 1] \) and \( \sigma^2 > 0 \), the following function is a spatio-temporal covariance on \( \mathbb{R}^d \times \mathbb{R} \) (separable if \( \beta = 0 \)):

\[
C(h, u) = \frac{\sigma^2}{(a|u|^{2\alpha} + 1)^{\beta d/2}} \exp\left( -\frac{c\|h\|^{2\gamma}}{(a|u|^{2\alpha} + 1)^{\beta \gamma}} \right) .
\]

(1.17)

We can then infer non-separable covariances using covariance mixtures (cf. Prop. 1.1): if \( \mu \) is a non-negative measure on some space \( W \) and for all \( w \in W \), \( C_S(\cdot, w) \) and \( C_T(\cdot, w) \) are stationary covariances for which:

\[
\int_W C_S(0, w)C_T(0, w)\mu(dw) < \infty,
\]

then (58; 147):

\[
C(h, u) = \int_W C_S(h, w)C_T(u, w)\mu(dw) < \infty
\]

is a stationary covariance function that is in general non-separable. For example,

\[
C(h, t) = \frac{\gamma^{n+1}}{(\|h\|^{\alpha} / a + |t|^{\beta} / b + \gamma)^{n+1}}, \quad 0 < \alpha, \beta \leq 2
\]

(1.18)

is a mixture of this type for a Gamma distribution with mean \( (n + 1) / \gamma \) and spatial and temporal covariances respectively proportional to \( \exp(-\|h\|^{\alpha} / a) \) and \( \exp(-|t|^{\beta} / b) \).

1.7 Spatial autoregressive models

Spatial autoregressive models (AR) are useful for analyzing, characterizing and interpreting real-valued spatial phenomena \( X = \{X_s, s \in S\} \) defined on discrete spatial networks \( S \) that have neighborhood geometry.

In domains like econometrics, geography, environmental studies and epidemiology, \( S \) does not have to be regular, because sites \( s \in S \) represent centers of geographical units dispersed in space and observations \( X_s \) denote the value of the variable of interest at the site \( s \). The irregularity of such networks \( S \) is an initial difference between spatial and temporal autoregressions, the latter usually having for \( S \) some interval in \( \mathbb{Z} \).

In other domains such as imaging, radiography and remote sensing, \( S \) is indeed regular, typically being a subset of \( \mathbb{Z}^d \). This property allows us to define stationary models and to bring the study of AR random fields on \( \mathbb{Z}^d \) closer to that of time series on \( \mathbb{Z}^1 \). There is nevertheless a fundamental difference: spatial models are inherently non-causal in the sense that, unlike time series, they are not defined with
respect to some order relation on $S$. While the use of temporal causality is entirely justified when modeling variables $X_t$ such as inflation rate, stock price and rate of river flow, this is not so in the spatial case where autoregressive dependency exists in all directions in space. For example, presence/absence of a plant in some parcel of land can depend on the presence/absence of the plant in neighboring parcels, in all directions.

We begin by rapidly presently the stationary models known as MA, ARMA and AR on $\mathbb{Z}^d$ (cf. (96) for a more in-depth treatment). After this, we consider ARs on finite general networks and in particular two important classes of AR: SAR (S for Simultaneous AR) and CAR (C for Conditional AR) models.

### 1.7.1 Stationary MA and ARMA models

#### MA models

Let $(c_s, s \in \mathbb{Z}^d)$ be a sequence in $l^2(\mathbb{Z}^d)$ (i.e., satisfying: $\sum_{s \in \mathbb{Z}^d} c_s^2 < \infty$) and $\eta$ a SWN on $\mathbb{Z}^d$ with variance $\sigma^2$. An MA($\infty$) model on $\mathbb{Z}^d$ (MA for Moving Average) is a linear process defined in $L^2$ by:

$$X_t = \sum_{s \in \mathbb{Z}^d} c_s \eta_{t-s}. \quad (1.19)$$

$X$ is the **infinite moving average** of the noise $\eta$ with respect to weights $c$.

**Proposition 1.5.** The covariance and spectral density of the MA process (1.19) on $\mathbb{Z}^d$ are respectively:

$$C(h) = \sigma^2 \sum_{t \in \mathbb{Z}^d} c_t c_{t+h} \text{ and } f(u) = \frac{\sigma^2}{(2\pi)^d} \left| \sum_{t \in \mathbb{Z}^d} c_t e^{it\cdot u} \right|^2.$$

**Proof.** $C$ can be calculated using bilinearity of covariances and the fact that $\eta$ is a SWN. As for the spectral density, it can be found using the Fourier inversion formula (1.4):

$$f(u) = \frac{\sigma^2}{(2\pi)^d} \sum_{h \in \mathbb{Z}^d} \sum_{t \in \mathbb{Z}^d} c_t c_{t+h} e^{it\cdot u} = \frac{\sigma^2}{(2\pi)^d} \left| \sum_{t \in \mathbb{Z}^d} c_t e^{it\cdot u} \right|^2.$$

We say we have an MA model if the support $M = \{s \in \mathbb{Z}^d : c_s \neq 0\}$ of the sequence of weights is finite. The covariance $C$ is zero outside of its support $S(C) = M - M = \{h : h = t - s \text{ for } s, t \in M\}$. Even though when $d = 1$ any covariance process with finite support has an MA representation, this is not the case when $d \geq 2$ (cf. Prop. 1.8).
ARMA models

These models are an extension of temporal \((d = 1)\) ARMA models: let \(P\) and \(Q\) be the following two polynomials of the \(d\)-dimensional complex variable \(z \in \mathbb{C}^d\),

\[
P(z) = 1 - \sum_{s \in R} a_s z^s \quad \text{and} \quad Q(z) = 1 + \sum_{s \in M} c_s z^s,
\]

with \(R\) (resp. \(M\)) the support of the AR (resp. MA) being finite subsets of \(\mathbb{Z}^d\) not containing the origin and for \(s = (s_1, s_2, \ldots, s_d)\), \(z^s = z_1^{s_1} \ldots z_d^{s_d}\). Let \(B^sX_t = X_{t-s}\) denote the \(s\)-translation operator in \(L^2\). In formal terms, an ARMA is associated with polynomials \(P\) and \(Q\) and an SWN \(\eta\) in \(L^2\) by:

\[
\forall t \in \mathbb{Z}^d : P(B)X_t = Q(B)\eta_t,
\]

or alternatively,

\[
\forall t \in \mathbb{Z}^d : X_t = \sum_{s \in R} a_s X_{t-s} + \eta_t + \sum_{s \in M} c_s \eta_{t-s}.
\]

Let \(\mathbb{T} = \{ \xi \in \mathbb{C}, |\xi| = 1 \}\) be the 1-dimensional torus. We have the following existence result:

**Proposition 1.6.** Suppose that \(P\) has no zeros on the torus \(\mathbb{T}^d\). Then equation (1.20) has a stationary solution \(X\) in \(L^2\). Denoting \(e^{iu} = (e^{iu_1}, \ldots, e^{iu_d})\), the spectral density of \(X\) is:

\[
f(u) = \frac{\sigma^2}{(2\pi)^d} \left| \frac{Q}{P} (e^{iu}) \right|^2,
\]

and its covariance is given by the Fourier coefficients of \(f\).

**Proof.** As \(P\) has no zeros on the torus \(\mathbb{T}^d\), \(P^{-1}Q\) has Laurent series development,

\[
P^{-1}(z)Q(z) = \sum_{s \in \mathbb{Z}^d} c_s z^s,
\]

which converges in a neighborhood of the torus \(\mathbb{T}^d\) and whose coefficients \((c_s)\) decrease exponentially fast to 0. This ensures that \(X_t = \sum_{s \in \mathbb{Z}^d} c_s \eta_{t-s}\) exists in \(L^2\), that it satisfies (1.20) and has spectral density

\[
f(u) = \frac{\sigma^2}{(2\pi)^d} \left| \sum_{s \in \mathbb{Z}^d} c_s e^{isu} \right|^2 = \frac{\sigma^2}{(2\pi)^d} \left| \frac{Q}{P} (e^{iu}) \right|^2.
\]

\(\Box\)
MA models correspond to the choice $P \equiv 1$ and AR models to $Q \equiv 1$. As with time series, interest in ARMA models is due to the fact that they can get “close” to random fields that have continuous spectral densities: in effect, for any dimension $d$, the set of rational fractions is dense (e.g., for the sup norm) in the space of continuous functions on the torus $T^d$.

As the spectral density of ARMA processes is rational, its covariance decreases exponentially fast to zero in the limit. Here again, as with time series, after a certain rank the covariances satisfy the linear recurrence relationships known as the Yule-Walker equations. In $\mathbb{Z}$, these equations can be solved analytically and provide a tool to identify the ranges $R$ and $M$ of the AR and MA parts and estimate parameters $a$ and $c$. However, in dimension $d \geq 2$, the Yule-Walker equations cannot be solved analytically. Furthermore, unlike time series ARMA models generally do not have a finite unilateral (or causal) representation with respect to the lexicographic order when $d \geq 2$ (cf. (1.8)).

Even though there is no theoretical reason limiting their use (cf. for example (119)), the preceding remarks explain why, unlike for time series ARMA models are rarely used in spatial statistics.

Nevertheless we note that semi-causal spatio-temporal models (non-causal in space but causal in time) can turn out to be well adapted to studying spatial dynamics: STARMA (Spatio-Temporal ARMA) models, introduced by Pfeifer and Deutsch (76; 174) fall into this category (cf. also (48, §6.8)).

Two types of autoregressive model, SAR and CAR models are frequently used in spatial analyses. First, let us take a look at stationary models.

### 1.7.2 Stationary simultaneous autoregression

To simplify, suppose that $X$ is centered. Let $R$ be a finite subset of $\mathbb{Z}^d$ not containing the origin. A stationary SAR (Simultaneous AR) model relative to the SWN $\eta$ and with parameters $a = \{a_s, s \in R\}$ is:

$$X_t = \sum_{s \in R} a_s X_{t-s} + \eta_t. \quad (1.21)$$

$X_t$ is the weighted sum of the values $X_u$ at the $R$-neighbors of $t$ along with added noise $\eta_t$. $X$ exists if the characteristic polynomial $P$ of the autoregression has no zero on the torus $T^d$, where

$$P(e^{i\lambda}) = 1 - \sum_{s \in R} a_s e^{i\lambda s}.$$  

Equations (1.21) can be interpreted as a system of simultaneous AR equations with the usual econometrical meaning: $\{X_{t-s}, s \in R\}$ are “spatially lagged” endogenous variables influencing the response $X_t$ at $t$ with site $u$ having influence on $t$ when $t - u \in R$. This relation defines the oriented graph $\mathcal{R}$ of the SAR model. We now present some examples.
**Example 1.8.** Several SAR models

**Semi-causal Space × Time models**

$s \in \mathbb{Z}$ gives the spatial coordinate and $t \in \mathbb{N}$ the temporal one; an example of Markov dynamics at time $t$ and location $s$ is:

$$\forall t \in \mathbb{N} \text{ and } s \in \mathbb{Z}: X_{s,t} = \alpha X_{s,t-1} + \beta (X_{s-1,t} + X_{s+1,t}) + \epsilon_{s,t}.$$  

The temporal connection $(s,t - 1) \rightarrow (s,t)$ has a direction whereas instantaneous spatial links $(s,t) \leftrightarrow (s \pm 1, t)$ do not. The lexicographic causal representation of this SAR is infinite (cf. Fig. 1.6). More precisely, for $\alpha = \beta = \delta/(1 + \delta^2)$, this semi-causal model has the following infinite causal representation for the lexicographic order (defined by $(u,v) \preceq (s,t)$ if $v < t$ or if $v = t$ and $u \leq s$; (24)):

$$X_{s,t} = 2\delta X_{s-1,t} + \delta^2 X_{s-2,t} - \delta X_{s-1,t-1} + \delta(1 - \delta^2) \sum_{j \geq 0} \delta^j X_{s+j,t-1} + \epsilon_{s,t}.$$  

**Isotropic four nearest neighbor SAR models on $\mathbb{Z}^2$**

$$X_{s,t} = \alpha(X_{s-1,t} + X_{s+1,t} + X_{s,t-1} + X_{s,t+1}) + \epsilon_{s,t}.$$  

Here, $\mathcal{R}$ is a symmetric graph; $X$ exists if and only if

$$\forall \lambda, \mu \in [0, 2\pi], \ P(\lambda, \mu) = 1 - 2\alpha(\cos \lambda + \cos \mu) \neq 0,$$

ensuring that spectral density $f(\lambda, \mu) = \sigma^2 P(\lambda, \mu)^{-2} \in L^2(T^2)$; this condition is satisfied iff $|\alpha| < 1/4$.

**Factorizing SAR(1) models**

An example of a factorizing SAR on $\mathbb{Z}^2$ is

$$X_{s,t} = \alpha X_{s-1,t} + \beta X_{s,t-1} - \alpha \beta X_{s-1,t-1} + \epsilon_{s,t}, \ |\alpha| \text{ and } |\beta| < 1.$$  

(1.22)
Noting $B_1$ and $B_2$ the lag operators relative to coordinates $s$ and $t$, equation (1.22) can be written:

$$(1 - \alpha B_1)(1 - \beta B_2)X_{s,t} = \varepsilon_{s,t}.$$ 

We deduce that the covariance $c$ of $X$ is separable:

$$c(s - s', t - t') = \sigma^2 \alpha |s - s'| \beta |t - t'|,$$

where $\sigma^2 = \sigma^2(1 - \alpha^2)^{-1}(1 - \beta^2)^{-1}$ is the product of a 1-dimensional AR(1) covariance with parameter $\alpha$ and a covariance of the same type with parameter $\beta$.

It is easy to generalize these models to autoregressions of any order $p = (p_1, p_2)$ and any dimension $d \geq 2$. Being able to factorize the AR polynomial and the covariance makes these models easy to work with (cf. §1.6).

SAR models are frequently used for their simplicity and because they involve few parameters. However, the following two problems must be dealt with:

1. Without imposing constraints, SAR models are not (in general) identifiable. Recall that we say model $M(\theta)$ is identifiable if the distributions it defines for two different $\theta$ are different; for example, in $\mathbb{Z}$, it is simple to show that the following three SAR models:

   (i) $X_t = aX_{t-1} + bX_{t+1} + \eta_t, \ t \in \mathbb{Z}$, $a \neq b$, $|a|, |b| < 1/2$,
   (ii) $X_t = bX_{t-1} + aX_{t+1} + \eta^*_t$ and
   (iii) $X_t = a_1X_{t-1} + a_2X_{t-2} + \varepsilon_t$

are identical for appropriate choices of $a_1, a_2$ and variances of the WN errors $\eta, \eta^*$ and $\varepsilon$ (it suffices to identify each of their spectral densities and to realize that we can impose constraints allowing us to make equal the three of them). This said, if we impose the constraint $a < b$, the model becomes identifiable.

2. As with simultaneous equations in econometrics, estimation of SAR models by ordinary least squares (OLS) on the residuals is not consistent (cf. Prop. 5.6).

### 1.7.3 Stationary conditional autoregression

Suppose $X$ is a centered stationary second-order process on $\mathbb{Z}^d$ with spectral density $f$. If $f^{-1} \in L^1(\mathbb{T}^d)$, $X$ has the infinite non-causal linear representation (96, Th. 1.2.2):

$$X_t = \sum_{s \in \mathbb{Z}^d \setminus \{0\}} c_s X_{t-s} + \varepsilon_t.$$ 

In this form, $c_s = c_{s-s}$ for all $s$ and $\varepsilon_t$ is a conditional residual, i.e., for any $s \neq t$, $\varepsilon_t$ and $X_s$ are uncorrelated.

This leads us to the following definition of an $L$-Markov random field or CAR(L) model: let $L$ be a finite symmetric subset of $\mathbb{Z}^d$ not containing the origin 0 and $L^+$ the positive half-space of $L$ with respect to the lexicographic order on $\mathbb{Z}^d$. 

Definition 1.9. A stationary CAR(L) model in $L^2$ is given by, for any $t \in \mathbb{Z}^d$,

$$X_t = \sum_{s \in L} c_s X_{t-s} + e_t \quad \text{with, if } s \in L^+: c_s = c_{-s};$$  \hfill (1.23)

$$\forall s \neq t : \text{Cov}(e_t, X_s) = 0 \text{ and } E(e_t) = 0.$$  

The absence of correlation between $X_s$ and $e_t$ when $s \neq t$ translates the fact that

$$\sum_{s \in L} c_s X_{t-s} - \sum_{s \in L^+} c_s (X_{t-s} + X_{t+s})$$  

is the best linear prediction in $L^2$ of $X_t$ using all other variables $\{X_s, s \neq t\}$; here, $X$ is an $L$-Markov random field in the linear prediction sense. If $X$ is a Gaussian random field, it is the best possible prediction and we say that $X$ is an $L$-Markov Gaussian process. CAR and SAR models differ in several respects:

1. CAR models require parametric constraints: $L$ must be symmetric and for all $s$, $c_s = c_{-s}$.
2. The conditional residuals $e_t$ do not represent a white noise. We say that $\{e_t\}$ is a colored noise.
3. Residuals $e_t$ are uncorrelated with $X_s$ when $s \neq t$.

Proposition 1.7. The model $X$ defined in (1.23) exists in $L^2$ if the characteristic polynomial of the CAR models,

$$P^*(\lambda) = (1 - 2 \sum_{s \in L^+} c_s \cos(\langle s, u \rangle))$$

is non-zero on the torus $\mathbb{T}^d$. In this case, the spectral density is

$$f_X(u) = \sigma^2_e (2\pi)^{-d} P^*(\lambda)^{-1}$$

and the conditional residuals form a correlated noise with covariance:

$$\text{Cov}(e_t, e_{t+s}) = \begin{cases} 
\sigma^2_e & \text{if } s = 0, \\
-\sigma^2 e c_s & \text{if } s \in L \text{ and } \text{Cov}(e_t, e_{t+s}) = 0 \text{ otherwise.}
\end{cases}$$

Proof. We first remark that $E(e_0 X_u) = 0$ if $u \neq 0$ and $E(e_0 X_0) = \sigma^2_e$ when $u = 0$. Since $e_0 = X_0 - \sum_{s \in L} c_s X_{-s}$, this orthogonality becomes, in the frequency domain:

$$\forall u \neq 0 : \int_{\mathbb{T}^d} e^{-i\langle \lambda, u \rangle} \left[1 - \sum_{s \in L} c_s e^{-i\langle \lambda, s \rangle}\right] f_X(\lambda) d\lambda = 0.$$

Plancherel’s theorem implies that $f_X(u) = \sigma^2_e (2\pi)^{-d} P^*(u)^{-1}$. As residual $e_t = X_t - \sum_{s \in L} c_s X_{t-s}$ is a linear filter spectral density of $X$, it has spectral density

$$f_e(u) = \sigma^2_e (2\pi)^{-d} P(u)^{-1} |P(u)|^2 = \sigma^2_e (2\pi)^{-d} P(u).$$

The result is thus proved. \qed
Note that whereas the spectral density of a CAR model is proportional to $P^*(u)^{-1}$, that of a SAR model is proportional to $|P(u)|^{-2}$. In dimension $d \geq 3$, the condition “$P^*$ has no zeros on the torus” is not necessary (cf. Ex. 1.12).

As for SAR models, the Yule-Walker equations for covariance of CAR models can be obtained by multiplying the equation defining $X_t$ by $X_s$ and then taking the expectation: for example, for an isotropic four nearest neighbor (4-NN) CAR model in $\mathbb{Z}^2$, these equations are:

$$\forall s: r(s) = \sigma^2 \delta_0(s) + a \sum_{t : \|t-s\|_1 = 1} r(t).$$

There are three reasons justifying modeling using CAR models:

1. CAR representations are intrinsic: they give the best linear prediction of $X_t$ from its other values $\{X_s, s \neq t\}$.
2. Estimating CAR models using OLS is consistent (cf. Prop. 5.6).
3. The family of stationary CAR models contains that of the SAR models, strictly so when $d \geq 2$.

**Proposition 1.8. Stationary SAR and CAR models on $\mathbb{Z}^d$.**

1. Every SAR model is a CAR model. In $\mathbb{Z}$, both classes are identical.
2. When $d \geq 2$, the family of CAR models is larger than that of the SAR models.

**Proof.**

1. To get the CAR representation of a SAR model: $P(B)X_t = \varepsilon_t$, we write the spectral density $f$ of $X$ and see that it is the spectral density of a CAR model by expanding $|P(e^{i\lambda})|^2 = |1 - \sum_{s \in R} a_s e^{i\lambda \cdot s}|^2$. We thus obtain the support $L$ of the CAR model and its coefficients after imposing the normalization $c_0 = 0$:

$$f(u) = \frac{\sigma^2}{(2\pi)^d |P(e^{i\lambda})|^2} = \frac{\sigma^2}{(2\pi)^d C(e^{i\lambda})}, \text{ with } c_0 = 1.$$

For $A - B = \{i - j : i \in A \text{ and } j \in B\}$, we get:

$$L = \{R^* - R^*\} \setminus \{0\}, \text{ where } R^* = R \cup \{0\} \text{ and }$$

$$\text{if } s \in L, c_s = (\sigma^2 / \sigma^2) \sum_{v, y + s \in R} a_v a_{y + s} \text{ if } s \neq 0 \text{ and } 0 \text{ otherwise.}$$

When $d = 1$ the SAR and CAR classes are identical due to Fejer’s theorem which states that any trigonometric polynomial $P^*(e^{i\lambda})$ of one complex variable for which $P^* \geq 0$ is the square modulus of a trigonometric polynomial: if $P^*(e^{i\lambda}) \geq 0$, $\exists P$ such that $P^*(e^{i\lambda}) = |P(e^{i\lambda})|^2$. Thus the CAR-$P^*$ model can be equated with the SAR-$P$ one.

2. We show that over $\mathbb{Z}^2$ the CAR model $X_t = c \sum_{s : \|s-t\|_1 = 1} X_s + \varepsilon_t$, $c \neq 0$ has no SAR representation. The spectral density of $X$ satisfies:

$$f_X^{-1}(\lambda_1, \lambda_2) = c(1 - 2c(\cos \lambda_1 + \cos \lambda_2)). \quad (1.24)$$
If some SAR had spectral density $f_X$, its support $R$ would satisfy $R \subseteq L$. Noting $(a_s)$ the coefficients of the SAR, we must either have $a_{(1,0)} \neq 0$ or $a_{(-1,0)} \neq 0$, say $a_{(1,0)} \neq 0$; similarly, either $a_{(0,1)}$ or $a_{(0,-1)} \neq 0$, say $a_{(0,1)} \neq 0$. This implies that a non-zero term that depends on $\cos(\lambda_1 - \lambda_2)$ has to appear in $f_X^{-1}$, which is not the case. Thus, CAR model (1.24) has no SAR representation.

MA processes with finite support have covariances with bounded range. When $d = 1$, Fejer’s theorem ensures that the converse is true: processes on $\mathbb{Z}$ having covariances with bounded range are MAs. This is no longer true for $d \geq 2$: for example, the random field with correlation $\rho$ at distance 1 and 0 at distances $> 1$ has no MA representation; this can be proved using similar arguments to those in part (2) of the previous proposition.

Let us present several examples of CAR representations of SAR models on $\mathbb{Z}^2$.

**Example 1.9.** SAR $\to$ CAR mappings

1. The causal AR (cf. Fig. 1.7-a) with support $R = \{(1,0), (0,1)\}$:

   $$X_{s,t} = \alpha X_{s-1,t} + \beta X_{s,t-1} + \epsilon_{s,t}$$

   is a CAR($L$) model with half-support $L^+ = \{(1,0), (0,1), (-1,1)\}$ and coefficients $c_{1,0} = \alpha \kappa^2$, $c_{0,1} = \beta \kappa^2$, $c_{-1,1} = -\alpha \beta \kappa^2$ and $\sigma^2_e = \kappa^2 \sigma^2_e$, where $\kappa^2 = (1 + \alpha^2 + \beta^2)^{-1}$.

2. The non-causal SAR model:

   $$X_{s,t} = a(X_{s-1,t} + X_{s+1,t}) + b(X_{s,t-1} + X_{s,t+1}) + \epsilon_{s,t}$$

   is a CAR($L$) model (cf. Fig. 1.7-b) with half-support $L^+ = \{(1,0), (2,0), (-1,1), (0,1), (0,2), (1,1), (0,2)\}$ and coefficients:

   $$c_{1,0} = 2a \kappa^2, \quad c_{0,1} = 2b \kappa^2, \quad c_{2,0} = 2a^2 \kappa^2, \quad c_{0,2} = 2b^2 \kappa^2$$

   $$c_{-1,1} = -2ab \kappa^2, \quad \sigma^2_e = \sigma^2_e \kappa^2 \quad \text{where} \quad \kappa^2 = (1 + 2a^2 + 2b^2)^{-1}.$$

3. The factorizing SAR model:

   $$X_{s,t} = \alpha X_{s-1,t} + \beta X_{s,t-1} - \alpha \beta X_{s-1,t-1} + \epsilon_{s,t}, \quad |\alpha| \quad \text{and} \quad |\beta| < 1$$

   is an 8-NN CAR model with coefficients:

   $$c_{1,0} = \alpha(1 + \alpha^2)^{-1}, \quad c_{0,1} = \beta(1 + \beta^2)^{-1}, \quad c_{1,1} = c_{-1,1} = -c_{1,0} \times c_{0,1}$$

   $$\sigma^2_e = \sigma^2_e \kappa^2, \quad \text{where} \quad \kappa^2 = (1 + \alpha^2)^{-1}(1 + \beta^2)^{-1}.$$

In these three examples, $\kappa^2 < 1$ is the gain in variance of the CAR prediction of $X$ with respect to the SAR prediction.
1.7.4 Non-stationary autoregressive models on finite networks $S$

A real-valued process on $S = \{1, 2, \ldots, n\}$ is a vectorial r.v. $X^* \in \mathbb{R}^n$. Non-stationarity of $X^*$ can influence the vector of expectations $\mu = E(X^*)$, the network $S$ and the covariance matrix $\Sigma = Cov(X^*)$. We only deal with second-order non-stationarity here, working with the centered process $X = X^* - \mu$ for which $\Sigma = Cov(X^*) = Cov(X)$.

Let $\epsilon = (\epsilon_t, t \in S)$ be a centered noise in $L^2$. MA, AR and ARMA representations of $X$, either site by site or in matrix notation in the basis $\epsilon$ are defined by the equations:

$$
MA : \quad X_t = \sum_{s \in S} b_{t,s} \epsilon_s, \text{ or } X = B \epsilon,
$$

$$
AR : \quad X_t = \sum_{s \in S, s \neq t} a_{t,s} X_s + \epsilon_t, \text{ or } AX = \epsilon,
$$

$$
ARMA : \quad X_t = \sum_{s \in S, s \neq t} a_{t,s} X_s + \sum_{s \in S} b_{t,s} \epsilon_s, \text{ or } AX = B \epsilon,
$$

where, for $s, t \in S$, we have $B_{t,s} = b_{t,s}$, $A_{t,s} = 1$, $A_{t,s} = -a_{t,s}$ when $t \neq s$. The MA representation always exists; the AR and ARMA ones do too as long as $A$ is invertible. Noting $\Gamma = Cov(\epsilon)$, these models are second-order characterized by their covariances $\Sigma$:
Suppose we choose \( \varepsilon \) to be a WWN with variance 1 \( (\Gamma = I_n) \) and denote \(<\) an arbitrary total order enumerating the points of \( S \). If \( X \) is centered with invertible covariance \( \Sigma \), then \( X \) has a unique causal AR representation relative to \( \varepsilon \) and order \(<\); this representation is associated with the lower triangular matrix \( A^* \) from the Cholesky factorization \( \Sigma = A^*A^* \). The fact that \( A^* \), like \( \Sigma \), depends on \( n(n+1)/2 \) parameters confirms identifiability of the causal AR model. Equivalent AR representations, generally non-identifiable are written \( \tilde{A}X = \eta \), where for some orthogonal matrix \( P \), \( \eta = P\varepsilon \) (\( \eta \) still a WWN with variance 1) and \( \tilde{A} = PA^* \).

In practice, AR models are associated with not necessarily symmetric influence graphs \( \mathcal{R} \): \( s \to t \) is a (directed) edge of \( \mathcal{R} \) if \( X_s \) influences \( X_t \) with some weight \( a_{t,s} \) and the neighborhood of \( t \) is defined as \( \mathcal{N}_t = \{ s \in S : s \to t \} \).

**Local one-parameter SAR representation**

Let \( W = (w_{t,s})_{t,s \in S} \) be a weights matrix or influence graph measuring the influence of \( s \) on \( t \) where, for each \( t \), \( w_{t,t} = 0 \); for example, \( W \) could be the spatial contiguity matrix made up of ones where \( s \) has influence over \( t \) and zeros elsewhere. Other choices of \( W \) are presented in Cliff and Ord’s book (45) (cf. §5.2 also). Once \( W \) has been chosen, a classical choice of spatial model for econometrics or spatial epidemiology is a SAR with parameter \( \rho \). If \( t \in S \) and if \( \varepsilon \) is a SWN(\( \sigma^2 \varepsilon \)),

\[
X_t = \rho \sum_{s:s \neq t} w_{t,s} X_s + \varepsilon_t, \text{ or } X = \rho WX + \varepsilon.
\]

This model is well-defined as long as \( A = I - \rho W \) is invertible.

**Markov CAR representation**

Once again, consider the centered vector \( X \). CAR representations are written in terms of *linear conditional expectation* (conditional expectation if \( X \) is a Gaussian random field):

\[
X_t = \sum_{s:s \neq t} c_{t,s} X_s + \varepsilon_t, \quad \forall t \in S,
\]

(1.25)

with \( E(\varepsilon_t) = 0 \), \( \text{Var}(\varepsilon_t) = \sigma_t^2 > 0 \) and \( \text{Cov}(X_t, \varepsilon_s) = 0 \) when \( t \neq s \). In this *intrinsic* representation, \( \varepsilon \) is a *conditional residual*.

CAR representations are associated with a neighborhood graph \( \mathcal{G} \) of \( S \) defined in the following way: \( s \to t \) is an edge of \( \mathcal{G} \) if \( c_{t,s} \neq 0 \). As we will see, \( \mathcal{G} \) is symmetric. Denote \( C \) the matrix with coefficients \( C_{s,s} = 0 \) and \( C_{t,s} = c_{t,s} \) when \( s \neq t \) and let \( D \) be the diagonal matrix with coefficients \( D_{t,t} = \sigma_t^2 \). The Yule-Walker equations
\[ \Sigma = C \Sigma + D \] can be obtained by multiplying (1.25) by \( X_s \) for \( s \in S \), then taking the expectation. \( \Sigma \) then satisfies:

\[ (I - C) \Sigma = D. \]

Hence, (1.25) defines a CAR model with regular covariance matrix \( \Sigma \) iff \( \Sigma^{-1} = D^{-1}(I - C) \) is symmetric and positive definite. In particular, representation (1.25) has to satisfy the constraints:

\[ c_{t,s} \sigma_s^2 = c_{s,t} \sigma_t^2, \quad \forall t \neq s \in S. \]  

(1.26)

Hence, \( c_{t,s} \neq 0 \) when \( c_{s,t} \neq 0 \), implying that the CAR’s graph \( G \) is symmetric. For algorithms that estimate CAR models, it is necessary to include these constraints. If \( X \) is stationary (for example on the finite torus \( S = (\mathbb{Z}/p\mathbb{Z})^d \)), we can reparametrize the model with \( c_{t-s} = c_{t,s} = c_{s-t} = c_{s,t} \) for \( t \neq s \). Under Gaussian hypotheses, (1.25) entirely characterizes this model.

We note that unlike stationary models on \( \mathbb{Z}^d \) (cf. Prop. 1.8), when \( S \) is finite the family of SAR models is the same as the CAR one.

**Markov Gaussian random fields**

Suppose \( X \) is a Gaussian process on \( S \), \( X \sim \mathcal{N}_n(\mu, \Sigma) \) with invertible \( \Sigma \) and where \( S \) is associated with a symmetric graph \( G \) without loops. Let \( \langle s, t \rangle \) mean \( s \) and \( t \) are neighbors in \( G \). We say that \( X \) is a \( G \)-Markov random field if, noting \( Q = (q_{s,t}) = \Sigma^{-1} \), \( q_{st} = 0 \) except when \( \langle s, t \rangle \). In this case, we have for all \( t \in S \),

\[ \mathcal{L}(\mathcal{L}(X_t X_s, s \neq t) \sim \mathcal{N}(\mu_t - q_{t,t}^{-1} \sum_{s: \langle t,s \rangle} q_{t,s} (X_s - \mu_s), q_{t,t}^{-1}) \]

and \( X \) follows a CAR model: for all \( t \in S \),

\[ X_t - \mu_t = -q_{t,t}^{-1} \sum_{s: \langle t,s \rangle} q_{t,s} (X_s - \mu_s) + e_t, \quad \text{Var}(e_t) = q_{t,t}^{-1}. \]  

(1.27)

Let \([Q]\) be the \( n \times n \) matrix with diagonal 0 and \([Q]_{t,s} = q_{t,s} \) if \( t \neq s \), and \( \text{Diag}(Q) \) the diagonal matrix whose diagonal is the same as that of \( Q \). (1.27) can be written

\[ X - \mu = -(\text{Diag})^{-1}[Q](X - \mu) + e. \]

As we will see in Chapter 2, Gaussian CAR models are Gibbs models with quadratic potentials (189).

**The Markov graph \( G \) of a SAR model**

Let \( \varepsilon \) be a Gaussian WN with variance \( \sigma^2 \). The Gaussian SAR model \( AX = \varepsilon \) exists if \( A^{-1} \) exists and has inverse covariance \( \Sigma^{-1} = Q = \sigma^{-2}(t AA) \) and SAR graph: \( \langle t, s \rangle \not\leftrightarrow a_{t,s} \neq 0 \). Its CAR representation (1.27) is:
1.7 Spatial autoregressive models

1. CAR coefficients are: \( c_{t,s} = -q_{t,s}/q_{t,t} \), where \( q_{t,s} = \sum_{l \in S} a_{t,l} a_{l,s} \);
2. The graph \( \mathcal{G} \) of the Markov CAR representation of \( X \) is:

\[
\langle t,s \rangle_{\mathcal{G}} \iff \begin{cases} 
\text{either } \langle t,s \rangle_{\mathcal{R}}, \\
\langle s,t \rangle_{\mathcal{G}}, \\
\text{or } \exists l \in S \text{ s.t. } \langle l,t \rangle_{\mathcal{R}} \text{ and } \langle l,s \rangle_{\mathcal{G}}.
\end{cases}
\]

\( \mathcal{G} \) is undirected with “double” the range of \( \mathcal{R} \) (cf. Fig. 1.8).

**Example 1.10.** CAR representation of nearest neighbor SAR models.

Let \( W = (w_{t,s})_{t,s \in S} \) be a weights matrix representing the influence of \( s \) on \( t \) with, for all \( t \), \( w_{t,t} = 0 \); consider the SAR with one parameter \( \rho \):

\[
X = \rho WX + \epsilon,
\]

where \( \epsilon \) is a \( \text{WN}(\sigma^2_\epsilon) \). The CAR model associated with this SAR model is given by (1.27) with \( \mu = 0 \) and precision matrix \( Q = \Sigma_X^{-1} \):

\[
Q = \sigma^{-2}_\epsilon (I - \rho(W + W^t) + \rho^2 WW).
\]

As for the best linear prediction of \( X \), it is given by the vector

\[
\hat{X} = -(\text{Diag})^{-1}QX.
\]

1.7.5 Autoregressive models with covariates

These types of models are especially used in spatial econometrics. Suppose that \( Z \) is a real-valued \( n \times p \) matrix of observable exogenous conditions. SARX models (X
for eXogenous) propose, in addition to regression of $X$ onto $Z$, a weights structure $W$ acting separately on the endogenous $X$ and exogenous $Z$ (7):

$$X = \rho WX + Z\beta + WZ\gamma + \epsilon, \quad \rho \in \mathbb{R}, \beta \text{ and } \gamma \in \mathbb{R}^p.$$  (1.28)

$X$ has three explicative factors: the usual regression variables ($Z\beta$), the endogenous ($\rho WX$) and spatially lagged exogenous variables ($WZ\gamma$) with the same weights vector $W$ but their own parameters.

A sub-model with common factors, also known as a spatial Durbin model is associated with the choice of constraint $\gamma = -\rho \beta$, i.e., with the regression model with SAR errors:

$$(I - \rho W)X = (I - \rho W)Z\beta + \epsilon \text{ or } X = Z\beta + (I - \rho W)^{-1}\epsilon.$$  (1.29)

The spatial lag sub-model corresponds to the choice $\gamma = 0$:

$$X = \rho WX + Z\beta + \epsilon.$$  (1.30)

Note that these models offer three different ways to model the mean: respectively,

$$E(X) = (I - \rho W)^{-1}[Z\beta + WZ\gamma], \quad E(X) = Z\beta \quad \text{and} \quad E(X) = (I - \rho W)^{-1}Z\beta,$$

but the same covariance structure $\Sigma^{-1} \times \sigma^2 = (I - \rho W')(I - \rho W)$ if $\epsilon$ is a WWN with variance $\sigma^2$. An estimation of these models using Gaussian maximum likelihood can be obtained by expressing the mean and variance of $X$ in terms of the model’s parameters.

Variants of these models are possible, for example by choosing $\epsilon$ to be a SAR model associated with weights matrix $H$ and some real-valued parameter $\alpha$. We can also let different weights matrices be associated with the endogenous and exogenous variables.

### 1.8 Spatial regression models

We talk of spatial regression when the process $X = (X_s, s \in S)$ is made up of a deterministic part $m(\cdot)$ representing large scale variations, drift, trend or mean of $X$, and $\epsilon$ a centered residuals process:

$$X_s = m(s) + \epsilon_s, \quad E(\epsilon_s) = 0.$$  

Depending on the context of the study and available exogenous information, there are many ways to model $m(\cdot)$, whether it be by regression (linear or otherwise), analysis of variance (qualitative exogenous variables), analysis of covariance (exogenous variables with quantitative and qualitative values) or with generalized linear models:
Response surface: \( m(s) = \sum_{l=1}^{p} \beta_l f_l(s) \) belongs to a linear space of known functions \( f_l \). If \( \{ f_l \} \) is a polynomial basis, the spanned space is invariant with respect to the coordinate origin. If \( s = (x, y) \in \mathbb{R}^2 \), a quadratic model in these coordinates is associated with the monomials \( \{ f_l \} = \{1, x, y, xy, x^2, y^2\}:
\[
m(x,y) = \mu + ax + by + cx^2 + dxy + ey^2.
\]

Exogenous dependency: \( m(s, z) = \sum_{l=1}^{p} \alpha_l z_s^{(l)} \) is expressed in terms of observable exogenous variables \( z_s \).

Analysis of variance: if \( s = (i, j) \in \{1, 2, \ldots, I\} \times \{1, 2, \ldots, J\} \), we consider an “additive” model: \( m(i, j) = \mu + \alpha_i + \beta_j \), where \( \sum_i \alpha_i = \sum_j \beta_j = 0 \).

Analysis of covariance: \( m(\cdot) \) is a combination of regressions and components of analysis of variance:
\[
m(s) = \mu + \alpha_i + \beta_j + \gamma z_s, \quad \text{with } s = (i, j).
\]

Cressie (48) suggested the following decomposition of the residuals \( \varepsilon_s \):
\[
X_s = m(s) + \varepsilon_s = m(s) + W_s + \eta_s + \varepsilon_s.
\]

\( W_s \) is a “smooth” component modeled by an intrinsic process whose range is in the order of \( c \) times \( (c < 1) \) the maximum distance between the observation sites; \( \eta_s \) is a micro-scale component independent of \( W_s \) with a range in the order of \( c^{-1} \) times the minimum distance between observation sites and \( \varepsilon_s \) is a measurement error or nugget component independent of \( W \) and \( \eta \).

Generally speaking, if \( X \) is observed at \( n \) sites \( s_i \in S \), linear models with linear spatial regressions are written:
\[
X_{si} = t z_{si} \delta + \varepsilon_{si}, \quad i = 1, \ldots, n,
\]
where \( z_{si} \) and \( \delta \in \mathbb{R}^p \), \( z_{si} \) is a covariate (qualitative, quantitative, mixed) observed at \( s_i \) and \( \varepsilon = (\varepsilon_{s_i}, i = 1, \ldots, n) \) spatially correlated centered residuals. Denoting \( X = t (X_{s_1}, \ldots, X_{s_n}) \), \( \varepsilon = t (\varepsilon_{s_1}, \ldots, \varepsilon_{s_n}) \) and \( Z = t (z_{s_1}, \ldots, z_{s_n}) \) the \( n \times p \) matrix of exogenous variables, (1.32) can be written in matrix form as:
\[
X = Z\delta + \varepsilon,
\]
with \( E(\varepsilon) = 0 \) and \( \text{Cov}(\varepsilon) = \Sigma \).

The second step consists of modeling \( \Sigma \) using a covariance function, variogram or spatial AR model.

Example 1.11. Rainfall in the State of Parana (parana dataset from the geoR package in R (181))

These data give the average rainfall during May-June over a number of years at 143 meteorological stations in the State of Parana, Brazil. The amount of rainfall can be influenced by various exogenous factors, climatic or otherwise, e.g., orography, though taking time-averages helps to diminish their effect. Upon examining
1 Second-order spatial models and geostatistics

the cloud of points shown in Figure 1.9-a, we notice that the phenomena is not, on average, stationary and suggest that an affine model of the response surface $m(s) = \beta_0 + \beta_1 x + \beta_2 y$, $s = (x,y) \in \mathbb{R}^2$ is a reasonable choice. It then remains to suggest a covariance on $\mathbb{R}^2$ for the residuals that would allow us to quantify the covariance $\Sigma$ of the 143 observations and validate the model in first and second-order (cf. §1.3.3).

Example 1.12. Modeling controlled experiments

The Mercer and Hall (156) dataset (cf. mercer-hall dataset on the website) gives the quantity of harvested wheat from an untreated field trial on a rectangular domain cut up into $20 \times 25$ parcels $(i,j)$ of the same size $2.5 \text{ m} \times 3.3 \text{ m}$. A first glance at Fig. 1.10-a showing amounts of harvested wheat does not easily help us determine whether the mean $m(\cdot)$ is constant or not. To try to come to a decision, we can use the fact that the data is on a grid to draw boxplots by row and column and attempt to discover if there is a trend (or not) in either direction.

A graphical analysis (cf. Fig. 1.10-b and 1.10-c) suggests that there is no trend with respect to rows $(i)$. We thus propose a model that only includes a column trend $(j)$:

$$X_{i,j} = \beta_j + \epsilon_{i,j}, \quad i = 1, \ldots, 20, \ j = 1, \ldots, 25.

Example 1.13. Percentage of gross agricultural produce consumed locally

Cliff and Ord (45) analyzed spatial variability of the percentage $X$ of gross agricultural output consumed where it was made in Ireland’s 26 counties $S$ (eire dataset in the spdep package). These percentages have been linked with an index $z$ measuring the quality of road access of each county (cf. Fig. 1.11-a). The dispersion diagram (Fig. 1.11-b) shows that the linear model,

$$X_s = \beta_0 + \beta_1 z_s + \epsilon_s, \quad s \in S,$n

is a reasonable choice. A preliminary analysis of the residuals of (1.33) estimated by OLS shows that there is spatial correlation in the data. We model this using a
weights matrix \( W = (w_{t,s})_{t,s \in S} \) with known weights representing the influence of \( s \) on \( t \). Figure 1.11-d shows the influence graph associated with the symmetric binary specification and we choose \( w_{t,s} = 1 \) if \( s \) and \( t \) are neighbors, \( w_{t,s} = 0 \) otherwise.

A first choice of model is the spatial lag model (1.30):

\[
X_s = \beta_0 + \beta_1 z_s + \gamma \sum_{t \in S} w_{s,t} X_t + \epsilon_s.
\]

A second possibility is to consider a regression with SAR residuals:

\[
X_s = \beta_0 + \beta_1 z_s + \epsilon_s, \quad \epsilon_s = \lambda \sum_{t \in S} w_{s,t} \epsilon_t + \eta_s,
\]

where \( \eta \) is a WWN. A model that generalized both choices is (cf. Ex. 1.21):
Fig. 1.11 (a) Percentage $X$ of gross agricultural produce consumed in each of the 26 counties of Ireland where it was produced; (b) road access index $Y$; (c) diagram showing dispersion between $X$ and $Y$; (d) influence graph associated with the binary specification.

$$X_s = \beta_0 + \beta_1 z_s + \gamma \sum_{t \in S} w_{s,t} X_t + \varepsilon_s,$$

where $\varepsilon_s = \lambda \sum_{t \in S} w_{s,t} \varepsilon_t + \eta_s$.

**1.9 Prediction when the covariance is known**

Our goal is to create a prediction map for $X$ over all $S$ when $X$ is only observed at a finite number of points of $S$. Kriging, introduced by Matheron, deals with this
prediction problem. It builds on the work of Krige (134), a South African mining engineer.

### 1.9 Simple kriging

Given \( n \) observations \( \{X_{s_1}, \ldots, X_{s_n}\} \) in \( S \), kriging aims to give a linear prediction of \( X_{s_0} \) at unobserved sites \( s_0 \). We suppose that the covariance (variogram) of \( X \) is known. If, as is the case in practice, it is not, it must be pre-estimated (cf. §5.1.3).

Denote:

\[
X_{s_0} = X_{s_0}, \quad X = t(X_{s_1}, \ldots, X_{s_n}), \quad \Sigma = \text{Cov}(X), \quad \sigma_0^2 = \text{Var}(X_0) \quad \text{and} \quad c = \text{Cov}(X, X_0), \quad c \in \mathbb{R}^n
\]

the second-order characteristics (known or estimated) of \( X \) and consider the linear predictor of \( X_{s_0} \):

\[
\hat{X}_0 = \sum_{i=1}^{n} \lambda_i X_{s_i} = t \lambda X.
\]

We keep the predictor that minimizes the mean of the square of errors \( e_0 = X_0 - \hat{X}_0 \) (MSE),

\[
\text{MSE}(s_0) = E\{(X_0 - \hat{X}_0)^2\}.
\]  \hspace{1cm} (1.34)

Simple kriging can be used when the mean \( m(\cdot) \) of \( X \) is known. Without loss of generality, we suppose that \( X \) is centered.

**Proposition 1.9.** Simple kriging: The linear predictor of \( X_{s_0} \) minimizing (1.34) and the variance of the prediction error are, respectively:

\[
\hat{X}_0 = t c \Sigma^{-1} X, \quad \tau^2(s_0) = \sigma_0^2 - t c \Sigma^{-1} c.
\]  \hspace{1cm} (1.35)

\( \hat{X}_0 \) is the Best Linear Unbiased Predictor (BLUP) of \( X_0 \), i.e., the one having the smallest mean square error.

**Proof.**

\[
\text{MSE}(s_0) = \sigma_0^2 - 2t \lambda c + t \lambda \Sigma \lambda = \Psi(\lambda);
\]

the minimum is located at some \( \lambda \) for which the partial derivatives of \( \Psi \) are zero. We find \( \lambda = \Sigma^{-1} c \) and it can be easily shown that it is a minimum. Substituting, we obtain the variance of the error given in (1.35).

**Remarks**

The optimal value of \( \lambda \) is none other than \( c = \Sigma \lambda \), i.e.,

\[
\text{Cov}(X_{s_i}, X_0 - \hat{X}_0) = 0 \quad \text{for} \quad i = 1, \ldots, n.
\]

These equations can be interpreted as showing that \( \hat{X}_0 \) is the orthogonal projection (with respect to the scalar product of the covariance of \( X \)) of \( X_0 \) onto the space
spanned by the variables $X_{s_1}, \ldots, X_{s_n}$. The predictor is identical to $X_0$ whenever $s_0$ is one of the observation sites. If $X$ is a Gaussian process, $\hat{X}_0$ is exactly the conditional expectation $E(X_0 \mid X_{s_1}, \ldots, X_{s_n})$; the distribution of this predictor is Gaussian and the error $X_0 - \hat{X}_0 \sim \mathcal{N}(0, \tau^2(s_0))$.

### 1.9.2 Universal kriging

More generally, suppose that $X = Z\delta + \varepsilon$ follows a spatial linear regression model (1.32). Given $z_0$ and the covariance $\Sigma$ of the residual $\varepsilon$ (but not the mean parameter $\delta$), we want to make an unbiased linear prediction of $X_0$, i.e., satisfying $\lambda^T X = t z_0$, which minimizes the mean square error (1.34). If $\Sigma$ is unknown, it first must be estimated (cf. §5.1.3).

**Proposition 1.10.** Universal kriging: the best unbiased linear predictor of $X_0$ is

$$\hat{X}_0 = \{ t^T c \Sigma^{-1} + (z_0 - t^T Z \Sigma^{-1} c)(t^T Z \Sigma^{-1} Z)^{-1} t^T Z \Sigma^{-1} \} X.$$  

(1.36)

The variance of the prediction error is

$$\tau^2(s_0) = \sigma_0^2 - t^T c \Sigma^{-1} + (z_0 - t^T Z \Sigma^{-1} c)(t^T Z \Sigma^{-1} Z)^{-1} (z_0 - t^T Z \Sigma^{-1} c).$$  

(1.37)

**Proof.** The MSE of the predictor $\lambda^T X$ is:

$$MSE(s_0) = \sigma_0^2 - 2 t^T \lambda c + t^T \lambda \Sigma \lambda.$$

We consider the quantity:

$$\phi(\lambda, \nu) = \sigma_0^2 - 2 \lambda^T c + \lambda \Sigma \lambda - 2 \nu (\lambda^T Z - t z_0),$$

where $\nu$ is a Lagrange multiplier. The minimum of $\phi$ is found where the partial derivatives of $\phi$ at $\lambda$ and $\nu$ are zero. For $\lambda$, we find $\lambda = \Sigma^{-1} (c + Z \nu)$. To obtain $\nu$, we substitute $\lambda$ into the unbiased condition and find

$$\nu = (t^T Z \Sigma^{-1} Z)^{-1} (z_0 - t^T Z \Sigma^{-1} c),$$

$$\lambda = \Sigma^{-1} c + \Sigma^{-1} Z (t^T Z \Sigma^{-1} Z)^{-1} (z_0 - t^T Z \Sigma^{-1} c).$$

By substitution into $MSE(s_0)$, we obtain (1.36) and (1.37).  

An interpretation of the universal kriging prediction (1.36) is as follows: we rewrite (1.36) as

$$\hat{X}_0 = t z_0 \hat{\delta} + c \Sigma^{-1} (X - Z \hat{\delta}),$$

where

$$\hat{\delta} = (t^T Z \Sigma^{-1} Z)^{-1} t^T Z \Sigma^{-1} X.$$
We will see in Chapter 5 that \( \hat{\delta} \) is the (optimal) generalized least squares (GLS) estimator of \( \delta \) (cf. §5.3.4). Universal kriging of \( X \) is thus the sum of the (optimal) estimation \( \hat{\delta}_{\text{GLS}} \) of \( E(X_0) = \hat{\delta}_{\text{GLS}} \delta \) and the simple kriging \( \hat{\delta}_{\text{GLS}} = \hat{\delta}_{\text{GLS}} + \hat{\delta}_{\text{res}} \) with residuals \( \hat{\delta}_{\text{res}} = (X - Z\hat{\delta}) \) estimated by GLS.

When \( X_s = m + \epsilon_s \) with unknown but constant \( m \), we say we are performing ordinary kriging.

Kriging formulae can be written analogously in terms of variograms if \( \epsilon \) is an intrinsic process (cf. Ex. 1.10); in effect, stationarity plays no part in obtaining results (1.36) and (1.37).

Kriging is an exact interpolator as \( \hat{X}_s = X_s \) if \( s \) is an observation site: in effect, if \( s_0 = s \) and if \( c \) is the \( i \)th column of \( \Sigma \), then \( \Sigma^{-1} c = \epsilon_i \) where \( \epsilon_i \) is the \( i \)th vector of the canonical basis of \( \mathbb{R}^n \) and \( Z\Sigma^{-1} c = \epsilon_0 \).

**Regularity of kriging surfaces**

Regularity at the origin of covariance \( C \) (variogram 2\( \gamma \)) determines the regularity of the kriging surface \( s \mapsto \hat{X}_s \), especially at the data sites (cf. Fig. 1.12):

1. For the nugget effect model, if \( s_0 \neq s_i \), then \( \hat{\delta} = n^{-1} \sum_{i=1}^{n} X_{s_i} \) and the prediction is none other than the arithmetic mean of the \( (X_{s_i}) \) if \( s_0 \neq s_i \), with peaks \( \hat{X}_{s_0} = X_{s_0} \) when \( s_0 = s_i \). More generally, if \( C(h) \) is discontinuous at 0, \( s \mapsto \hat{X}_s \) is discontinuous at the data sites.

2. If \( C(h) \) is linear at the origin, \( s \mapsto \hat{X}_s \) is everywhere continuous but not differentiable at the data sites.

3. If \( C(h) \) is parabolic at 0, \( s \mapsto \hat{X}_s \) is everywhere continuous and differentiable.

If in dimension \( d = 1 \) kriging is done with cubic covariance (1.8), the interpolation function is a cubic spline. In higher dimensions and for separable cubic covariances, predictions are piecewise cubic in each variable (43, p. 272). Laslett (140) gives empirical comparisons between spline functions and kriging predictions.

**Example 1.14.** Kriging the rainfall data for the State of Parana (continued).

After preliminary analyses, we estimate (cf. §5.3.4) the affine regression model \( m(s) = \beta_0 + \beta_1 x + \beta_2 y \), \( s = (x, y) \in \mathbb{R}^2 \) using a Gaussian covariance and a nugget effect for the residuals. Figure 1.13-a shows the prediction map using universal kriging with this model and Figure 1.13-b shows its standard error.

**1.9.3 Simulated experiments**

Simulated experiments are procedures aiming to learn a program (metamodel) \( y = f(x) \) that associates input \( x \in S = [0, 1]^d \) to output \( y \in \mathbb{R}^q \) (132; 193). Here, the “spatial” dimension \( d \) of input \( x \) is generally \( \geq 3 \). Existing spatial statistics methods
suppose that \( y \) is random and associated with \( x \) via a spatial model, for example a spatial regression

\[
y = \int z(x) \beta + \epsilon(x),
\]

where \( z(x) \) is known, \( \beta \in \mathbb{R}^p \) unknown and \( \epsilon \) is a stationary Gaussian process with covariance \( C \). Usually, we choose a separable \( C \) to have quick algorithms. If we have observations (calculations) \( x_i \mapsto y_i \) at points of some experimental design \( \Omega = \{x_1, x_2, \ldots, x_n\} \) of \( S \), universal kriging \( \hat{y}_\Omega = \{\hat{y}_\Omega(x), x \in S\} \) with covariance \( C \) gives a
Exercises

47

prediction of $y$ over all of $S$. If the covariance depends on some unknown parameter $\theta$, we must first pre-estimate $\theta$ and use universal kriging with $C_\theta$. For a fixed budget and given prediction criteria, our goal is to choose the optimal $\theta$ that minimizes the criteria over all $S$.

Choosing a sampling scheme for selecting observation sites

Suppose that $X$ is a random field with constant mean and that we want to choose an experimental design $\mathcal{O} = \{x_1, x_2, \ldots, x_n\}$ of $n$ points that minimizes the integral of the variance over a region of interest $A \subset \mathbb{R}^d$,

$$V(\mathcal{O}, A) = \int_A E[X_s - \hat{X}_s(\mathcal{O})]^2 ds = \int_A \tau^2(s, \mathcal{O}) ds. \tag{1.38}$$

In this set-up, $\hat{X}_s(\mathcal{O})$ and $\tau^2(s, \mathcal{O})$ are respectively the prediction and variance of the prediction using ordinary kriging. An approximation of (1.38) can be calculated by discretizing $A$ at a finite subset $R \subset A$ of cardinality $M > m$. Minimizing (1.38) over $R$ necessitates an exhaustive search over a set with $(M^m)$ elements. In practice, we use the following sequential algorithm:

1. Let $\mathcal{O}_{k-1} = \{x^*_1, \ldots, x^*_k\}$ be the first $k - 1$ chosen points and $R_k = R \setminus \mathcal{O}_{k-1}$.
   Choose $x^*_k = \text{arg min}_{x \in R_k} V(\mathcal{O}_{k-1} \cup s, A)$.
2. Repeat until $k = m$.

This $V$ criteria can be generalized to non-stationary mean processes. Other criteria may also be considered.

Exercises

1.1. Effect of spatial correlations on the variance of empirical means.

Suppose $X = \{X_s, s \in \mathbb{Z}^d\}$ is a stationary random field on $\mathbb{Z}^d$ with mean $m$ and covariance $C(h) = \sigma^2 \rho \|h\|_1$, where $\|h\|_1 = \sum_{i=1}^d |h_i|$ and $|\rho| < 1$.

1. For $d = 1$ and $\overline{X} = \sum_{t=1}^9 X_t / 9$, show that:

$$V_1(\rho) = \text{Var}\{\overline{X}\} = \frac{\sigma^2}{81} \{9 + 2 \sum_{k=1}^8 (9 - k) \rho^k\}.$$  

2. For $d = 2$ and $\overline{X} = \sum_{t=1}^3 \sum_{s=1}^3 X_{s,t} / 9$, show that:

$$V_2(\rho) = \text{Var}(\overline{X}) = \frac{\sigma^2}{81} \{9 + 24 \rho + 28 \rho^2 + 16 \rho^3 + 4 \rho^4\}.$$  

Compare $V_1(0), V_1(\rho)$ and $V_2(\rho)$. Show that for $\rho = 1/2$, these three values are respectively proportional to 9, 15.76 and 30.25.
3. Denote $N = n^d$ and $V_d = \text{Var}(X_N)$ where $X_N = N^{-1} \sum_{i \in \{1,2,\ldots,n\}^d} X_i$. Show that:

$$V_d(\rho) = \text{Var}(X_N) = \frac{\sigma^2}{N^2} \left\{ \frac{n(1+\rho - 2\rho^2 + o_n(1))}{1-\rho^2} \right\}^d.$$ 

Compare the width of the confidence intervals for $m$ under $\rho$-correlation and under independence.

1.2. Three methods of prediction for factorizing ARs on $\mathbb{Z}^2$.

Let $X$ be a stationary centered Gaussian process on $\mathbb{Z}^2$ with covariance $C(h) = \sigma^2 \rho^{|h|}$. Consider the following three predictions of $X_0$:

1. The predictor via $0 = E(X_0)$.
2. The optimal causal SAR predictor using $X_{1,0}$, $X_{0,1}$ and $X_{1,1}$.
3. The optimal CAR predictor using $\{X_i, t \neq 0\}$.

Give explicit representations of the last two predictors. Show that when $\rho = 1/2$, the variances of the prediction errors are respectively proportional to 1, 0.5625 and 0.36.

1.3. Krige’s formula.

Let $X = \{X_t, t \in \mathbb{R}^d\}$ be a centered stationary process in $L^2$ with covariance $C$. For a bounded Borel set $V \in \mathcal{B}(\mathbb{R}^d)$, note:

$$X(V) = \frac{1}{V(V)} \int_V X(z) dz \quad \text{and} \quad C(u, U) = \frac{1}{v(v)v(U)} \int_u \int_U C(y-z) dy dz,$$

where $u, U \in \mathcal{B}(\mathbb{R}^d)$ have volumes $v(u) > 0$ and $v(U) > 0$.

1. The extension variance of $X$ from $\nu$ to $V$ is defined as $\sigma^2_{E}(\nu, V) = \text{Var}(X(\nu) - X(V))$: i.e., it is the variance of the prediction error of $X(V)$ when using $\hat{X}(V) = X(\nu)$. Show that $\sigma_{E}^2(\nu, V) = C(\nu, \nu) + C(V, V) - 2C(\nu, V)$.

2. Suppose that $D \subset \mathbb{R}^d$ is partitioned into $I$ subdomains $V_i$ and each $V_i$ in turn is divided into $J$ equally sized subdomains $v_{ij}$ in such a way that we pass from partition $V_i$ to $V_j$ via a translation. We denote by $\nu$ the shared generating form of the $v_{ij}$ and $V$ that of the $V_i$. Noting $X_{ij} = X(v_{ij})$ and $X_i = \frac{1}{J} \sum_{j=1}^J X_{ij}$, we define the empirical dispersion variance and dispersion variance of $X$ for $\nu$ in $V$, $v \subset V$, by:

$$s^2(\nu \mid V) = \frac{1}{J} \sum_{j=1}^J (X_{ij} - X_i)^2 \quad \text{and} \quad \sigma^2(\nu \mid V) = E\{s^2(\nu \mid V)\}.$$

a. Show that $\sigma^2(\nu \mid V) = C(\nu, \nu) - C(V, V)$.

b. Show that $\sigma^2(\nu \mid D) = \sigma^2(\nu \mid V) + \sigma^2(V \mid D)$.

1.4. A sufficient p.s.d. condition for matrices.

1. Show that $C$ is p.s.d if $C$ is diagonally dominant, i.e., if for all $i: C_{ii} \geq \sum_{j \neq i} C_{ij}$.
2. Investigate this condition for the two covariances:
   a. $C(i, j) = \rho^{|i-j|}$ on $\mathbb{Z}$;
   b. $C(i, j) = 1$ if $i = j$, $\rho$ if $|i-j|_{\infty} = 1$ and 0 otherwise, on $\mathbb{Z}^d$.

Show that the converse of 1 is not true.

1.5. Product covariance, restriction and extension of covariances.

1. Show that if $C_k(\cdot)$ are stationary covariances on $\mathbb{R}^1$, the function $C(h) = \prod_{k=1}^d C_k(h_k)$ is a covariance on $\mathbb{R}^d$.

2. Show that if $C$ is a covariance on $\mathbb{R}^d$, then $C$ is a covariance on any vectorial subspace of $\mathbb{R}^d$.

3. Consider the function $C_0(h) = (1 - |h|/\sqrt{2}) \mathbf{1}\{|h| \leq \sqrt{2}\}$ on $\mathbb{R}$.

   a. Show that $C_0$ is a covariance on $\mathbb{R}$.
   b. For $(i, j) \in A = \{1, 2, \ldots, 7\}^2$, suppose $s_{ij} = (i, j)$ and $a_{ij} = (-1)^{i+j}$. Show that $\sum_{(i,j),(k,l) \in A} a_{ij} a_{kl} C_0(\|s_{ij} - s_{kl}\|) < 0$. Deduce that $C(h) = C_0(\|h\|)$ is not a covariance on $\mathbb{R}^2$.
   c. Show that $\sum_{u,v \in \{0,1\}^d} (-1)^{|u|+|v|} C_0(\|u-v\|) < 0$ when $d \geq 4$. Deduce that $C(h) = C_0(\|h\|)$ is not a covariance on $\mathbb{R}^d$ when $d \geq 4$.
   d. Show that $C(h) = C_0(\|h\|)$ is not a covariance on $\mathbb{R}^3$. Deduce that no isotropic extension of $C_0$ is a covariance on $\mathbb{R}^d$ if $d \geq 2$.

1.6. $\chi^2$ random fields.

1. If $(X,Y)$ is a pair of Gaussian variables, show that:

   \[
   \text{Cov}(X^2, Y^2) = 2\{\text{Cov}(X, Y)\}^2.
   \]

   Hint: if $(X,Y,Z,T)$ is a Gaussian vector in $\mathbb{R}^4$, then

   \[
   \]

2. Suppose that $X^1, \ldots, X^n$ are $n$ centered i.i.d. Gaussian processes on $\mathbb{R}^d$, each with covariance $C_X$. Show that the random field $Y$ defined by

   \[
   Y = \{ Y_s = \sum_{i=1}^n [X_s^i]^2, s \in \mathbb{R}^d \}
   \]

   is stationary with covariance $C_Y(h) = 2n C_X(h)^2$.

1.7. Markov property of exponential covariances.

Consider a stationary process $X$ on $\mathbb{R}$ with covariance $C(t) = \sigma^2 e^{-|t|}$, with $|\rho| < 1$. $X$ is observed at $n$ sites $\{s_1 < s_2 < \ldots < s_n\}$.

1. Show that if $s_0 < s_1$, the kriging of $X$ at $s_0$ is $\hat{X}_{s_0} = \rho^{s_1-s_0} X_{s_1}$.

2. Show that if $s_k < s_0 < s_{k+1}$, the kriging $\hat{X}_{s_0}$ only depends on $X_{s_k}$ and $X_{s_{k+1}}$ and give an explicit formulation of this kriging.
1.8. For a stationary process $X$ on $\mathbb{R}$ with covariance $C$, we observe: $X_0 = -1$ and $X_1 = 1$. Show that simple kriging gives

$$\hat{X}_s = \frac{C(s-1) - C(s)}{C(0) - C(1)}$$

and that the variance of the prediction error at $s$ is

$$\tau^2(s) = C(0) \left( 1 - \frac{(C(s) + C(s-1))^2}{C(0)^2 - C(1)^2} \right) + 2 \frac{C(s)C(s-1)}{C(0) - C(1)}.$$ 

Draw the graphs of $s \mapsto \hat{X}_s$ and $s \mapsto \tau^2(s)$ for $s \in [-3, 3]$ when $C$ is Matérn’s covariance with parameters $C(0) = 1, a = 1, \nu = 1/2$ (resp. $\nu = 3/2$). Comment on these graphs.

1.9. Models with zero correlation for distances $> 1$.

Consider the stationary process $X = (X_s, t \in \mathbb{Z})$ with correlation $\rho$ at a distance 1 and correlation 0 at distances $> 1$. Denote $X(n) = (X_1, X_2, \ldots, X_n)$ and $\Sigma_n = \text{Cov}(X(n))$.

1. Under what condition on $\rho$ is $\Sigma_3$ p.s.d.? Same question for $\Sigma_4$.

Find the correlation function of the MA(1) process: $X_t = \varepsilon_t + a \varepsilon_{t-1}$, where $\varepsilon$ is a SWN. Deduce that the condition $|\rho| \leq 1/2$ ensures that for all $n \geq 1$, $\Sigma_n$ is p.s.d.

2. Calculate the kriging of $X_0$ using $X_1$ and $X_2$. Same question when:

a. $E(X_t) = m$ is unknown.
b. $E(X_t) = at$.

3. Try to answer similar questions when $X = (X_s, (s, t) \in \mathbb{Z}^2)$ is a stationary random field with correlation $\rho$ at (Euclidean) distance 1 and 0 at distances $> 1$.

1.10. If $X_s = m + \varepsilon_s$, where $\{\varepsilon_s, s \in S\}$ is an intrinsic process with variogram $2\gamma$, give the kriging predictions (1.36) and kriging variances (1.37) as a function of $\gamma$.

1.11. Consider the process $X_s = \cos(U + sV), s \in \mathbb{R}$, where $U$ is the uniform distribution $\mathcal{U}(0, 2\pi)$, $V$ a Cauchy variable on $\mathbb{R}$ (with density $1/\pi(1 + x^2)$) and $U$ and $V$ independent. Show that $E(X_t) = 0$ and $\text{Cov}(X_t, X_t) = 2^{-1} \exp\{-|s-t|\}$. Deduce that the trajectories of $X$ are infinitely differentiable but that $X$ is not $L^2$ differentiable.


Show that the equation $X_t = \frac{1}{d} \sum_{s:||s-t||_1=1} X_s + \varepsilon_t$ defines a CAR model on $\mathbb{Z}^d$ if and only if $d \geq 3$.

1.13. Give explicit CAR representations of the following SAR models (graph, coefficients, variance ratio $\kappa^2$):

1. $X_{s,t} = aX_{s-1,t} + bX_{s,t-1} + cX_{s+1,t-1} + \varepsilon_{s,t}, (s, t) \in \mathbb{Z}^2$.
2. $X_{s,t} = a(X_{s-1,t} + X_{s+1,t}) + b(X_{s,t-1} + X_{s,t+1}) + c(X_{s-1,t-1} + X_{s+1,t+1}) + \varepsilon_{s,t}, (s, t) \in \mathbb{Z}^2$. 
Exercises 51

3. \(X_t = aX_{t-1} + bX_{t+2} + \varepsilon_t, (s,t) \in \mathbb{Z}\).

1.14. Simulating factorizing SAR models on \(\mathbb{Z}^2\).

Consider the factorizing centered Gaussian SAR model:

\[
X_{s,t} = \alpha X_{s-1,t} + \beta X_{s,t-1} - \alpha \beta X_{s-1,t-1} + \varepsilon_{s,t}, \quad 0 < \alpha, \beta < 1.
\]

If \(\text{Var}(\varepsilon) = \sigma^2 = (1 - \alpha^2)(1 - \beta^2)\), the covariance of \(X\) is \(C(s,t) = \alpha^{|s|} \beta^{|t|}\).

We propose three ways to simulate \(X\) on the rectangle \(S = \{0, 1, 2, \ldots, n-1\} \times \{0, 1, 2, \ldots, m-1\}\), of which the first and third are exact simulations:

1. Using the Cholesky decomposition of \(\Sigma = \Sigma_1 \otimes \Sigma_2\), give this decomposition and the associated simulation algorithm.
2. Using Gibbs sampling with the associated CAR model (cf. §4.2; give the associated CAR model and the simulation algorithm).
3. Using a recursive formulation that exactly defines \(X\) on \(S\): let \(W\) be the Brownian sheet on \((\mathbb{R}^+)^2\) (cf. (1.1)) and define variables

\[
Z_{s,0} = \alpha^s \times W([0, \alpha^{-2s}] \times [0, 1]) \quad \text{for } s = 0, \ldots, n-1 \quad \text{and}
\]

\[
Z_{0,t} = \beta^t \times W([0, 1] \times [0, \beta^{-2t}]) \quad \text{for } t = 0, \ldots, m-1.
\]

Show that \(Z\) has the same covariance as \(X\) in the directions of the two axes. Deduce an exact simulation method for \(X\) on a subgrid of \(S\).

1.15. The Yule-Walker equations.

Give the Yule-Walker equations for the stationary models with spectral density:

1. \(f(u_1, u_2) = \sigma^2(1 - 2a \cos u_1 - 2b \cos u_2)\).
2. \(g(u_1, u_2) = (1 + 2 \cos u_1)f(u_1, u_2)\).

1.16. Identify in series form the variance of the isotropic 4-NN CAR model on \(\mathbb{Z}^2\) with parameter \(a\), \(|a| < 1/4\). Determine the correlation at distance 1 of this model and draw the correlation graph \(a \rightarrow \rho(a)\). Same question in dimension \(d = 1\) with \(|a| < 1/2\).

1.17. Behavior of CAR models at their parametric boundary.

1. Consider the isotropic 4-NN CAR model on \(\mathbb{Z}^2\) with parameter \(a = 1/4 - \varepsilon\), with \(\varepsilon \downarrow 0\). What equation satisfies \(\rho_{\varepsilon} = \rho_X(1,0)\)? Show that \(1 - \rho_{\varepsilon} \sim -((\pi/2)(\log \varepsilon)^{-1})\) when \(\varepsilon\) is close to 0. For what values of \(a\) do we get \(\rho_{\varepsilon} = 0.9, 0.95\) and \(0.99\)?
2. Attempt the same questions for the (isotropic) 2-NN model on \(\mathbb{Z}^1\). Compare the behavior of \(\rho(\varepsilon)\) when \(d = 1\) and \(d = 2\) for small \(\varepsilon\).

1.18. The restriction of a Markov random field in \(\mathbb{Z}^2\) to \(\mathbb{Z}^1\).

Suppose \(X\) is the isotropic 4-NN CAR model on \(\mathbb{Z}^2\). What is the spectral density of the one parameter process \(\{X_{s,0}, s \in \mathbb{Z}\}\)? In \(\mathbb{Z}\), is this model still a Markov random field?
1.19. Exchangeable Gaussian models on \( S = \{1, 2, \ldots, n\} \).
Consider the following SAR model on \( \mathbb{R}^n \):

\[
X = \alpha J X + \varepsilon,
\]

where \( \varepsilon \) is a Gaussian WN and \( J \) the \( n \times n \) matrix with coefficients \( J_{ij} = 1 \) if \( i \neq j \), \( J_{ii} = 0 \) otherwise. We say that the set \( \{aI + bJ, a, b \in \mathbb{R}\} \) is stable under multiplication and inversion if \( aI + bJ \) is invertible.

1. Under what condition on \( \alpha \) is the model well-defined? Show that \( \text{Cov}(X) = r_0 I + r_1 J \) and find \( r_0 \) and \( r_1 \).
2. After identification of \( \beta \) and \( \text{Cov}(e) \), show that \( X \) can be written in the following CAR form:

\[
X = \beta J X + e.
\]

1.20. Suppose that \( X \) is a non-stationary Gaussian SAR model over the sites \( S = \{1, 2, \ldots, n\} \). Give its CAR representation: graph, joint distribution, conditional distributions and prediction of \( X_i \) using the other observations.

1.21. Two SARX models with covariates.
Let \( Y \) be an \( n \times p \) matrix of deterministic covariates that influence some spatial variable \( X \in \mathbb{R}^n \), \( W \) a spatial contiguity matrix on \( S = \{1, 2, \ldots, n\} \) and \( \eta \) a Gaussian WN with variance \( \sigma^2 \). Suppose we are interested in the following two models: first, the spatial lag model defined as the SAR model:

\[
X = Y\beta + \alpha WX + \eta
\]

with exogenous \( Y \); second, the Durbin model, defined as the SAR model on the residuals \( X - Y\beta = \alpha W(X - Y\beta) + \eta \). If \( I - \alpha W \) is invertible, calculate the distribution of \( X \) for each model. How do the results change if \( \eta \) is itself a SAR model, \( \eta = \rho \Delta \eta + e \), where \( \Delta \) is a known proximity matrix and \( e \) a Gaussian WN with variance \( \sigma^2_e \)? Calculate the log-likelihood of each model.
Spatial Statistics and Modeling
Gaetan, C.; Guyon, X.
2010, XIV, 302 p., Hardcover
ISBN: 978-0-387-92256-0