Springer Series in Statistics

Advisors:
P. Bickel, P. Diggle, S. Fienberg,
U. Gather, I. Olkin, S. Zeger

For other titles published in this series, go to
www.springer.com/series/692

Carlo Gaetan · Xavier Guyon

Spatial Statistics and Modeling

Translated by Kevin Bleakley

Springer

Carlo Gaetan
Dipartimento di Statistica
Università Ca' Foscari Venezia
San Giobbe - Cannaregio, 873
30121 Venezia VE
ltaly
gaetan@unive.it

Xavier Guyon
SAMOS
Université Paris I
90 rue de Tolbiac
75634 Paris CX 13
France
guyon@univ-paris1.fr

ISBN 978-0-387-92256-0 e-ISBN 978-0-387-92257-7
DOI 10.1007/978-0-387-92257-7
Springer New York Dordrecht Heidelberg London

Library of Congress Control Number: 2009938269

Preface

Spatial analysis methods have seen a rapid rise in popularity due to demand from a wide range of fields. These include, among others, biology, spatial economics, image processing, environmental and earth science, ecology, geography, epidemiology, agronomy, forestry and mineral prospection.

In spatial problems, observations come from a spatial process $X = \{X_s, s \in S\}$ indexed by a spatial set S, with X_s taking values in a state space E. The positions of observation sites $s \in S$ are either fixed in advance or random. Classically, S is a 2-dimensional subset, $S \subseteq \mathbb{R}^2$. However, it could also be 1-dimensional (chromatography, crop trials along rows) or a subset of $\mathbb{R}^3$ (mineral prospection, earth science, 3D imaging). Other fields such as Bayesian statistics and simulation may even require spaces S of dimension $d \geq 3$. The study of spatial dynamics adds a temporal dimension, for example $(s,t) \in \mathbb{R}^2 \times \mathbb{R}^+$ in the 2-dimensional case.

This multitude of situations and applications makes for a very rich subject. To illustrate, let us give a few examples of the three types of spatial data that will be studied in the book.

Geostatistical data

Here, S is a *continuous* subspace of $\mathbb{R}^d$ and the random field $\{X_s, s \in S\}$ observed at n fixed sites $\{s_1, \ldots, s_n\} \subset S$ takes values in a real-valued state space E. The rainfall data in Figure 0.1-a and soil porosity data in Fig. 0.1-b fall into this category. Observation sites may or may not be regularly spaced. Geostatistics tries to answer questions about modeling, identification and separation of small and large scale variations, prediction (or kriging) at unobserved sites and reconstruction of X across the whole space S.

Lattice data and data on fixed networks

Here, S is a *fixed discrete* non-random set, usually $S \subset \mathbb{R}^d$ and X is observed at points in S. Points s might be geographical regions represented as a network with

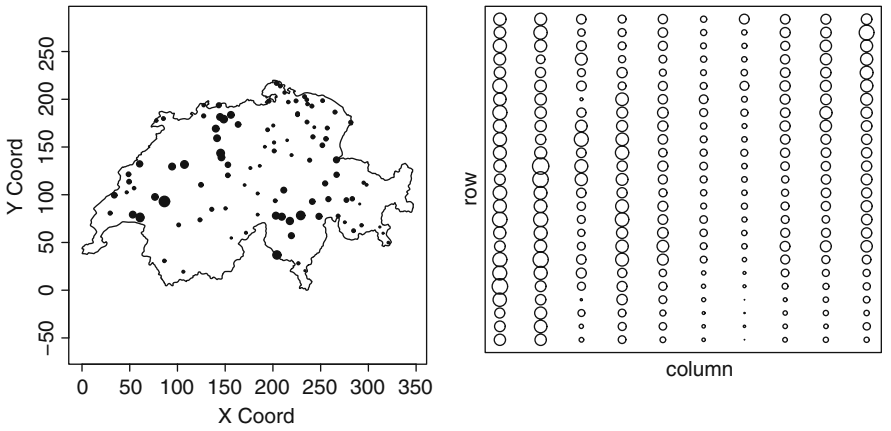

Fig. 0.1 (a) Rainfall over the Swiss meteorological network on May 8, 1986 (during the passage of Chernobyl's radioactive cloud. This is the `sic` dataset from the `geoR` package of *R* (178)); (b) Soil porosity (`soil` dataset from the `geoR` package). For both (a) and (b), the size of symbols are proportional to the value of X_s.

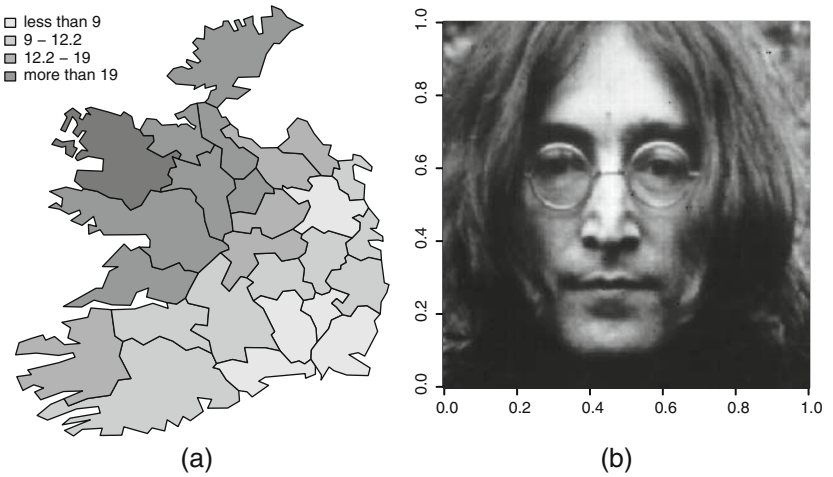

Fig. 0.2 (a) Percentage of people with blood group A in the 26 counties of Ireland (`eire` dataset from the `spdep` package); (b) Image of John Lennon (256×256 pixels in a 193-level grayscale, `lennon` dataset from the `fields` package).

given adjacency graph $\mathcal{G}$ (cf. the 26 counties of Ireland, Fig. 0.2-a) and X_s some value of interest measured at s. The state space E may or may not be real-valued. In image analysis, S is a regularly spaced set of pixels (cf. Fig. 0.2-b). Goals for these types of data include constructing and analyzing explicative models, quantifying spatial correlations, prediction and image restoration.

Point data

Figure 0.3-a shows the location of cell centers in a histological section seen under a microscope and Figure 0.3-b the location and size of pine trees in a forest. Here, the set of observation sites $x = \{x_1, x_2, \ldots, x_n\}$, $x_i \in S \subset \mathbb{R}^d$ is *random*, along with the number $n = n(x)$ of observation sites; x is the outcome of a spatial point process (PP) observed in window S. The process X is said to be *marked* if at each x_i we record a value, for example the diameter of the pine trees found at x_i. A central question in the statistical analysis of PPs is to know if the distribution of points is essentially regular (Figure 0.3-a), completely random (Poisson PP) or aggregated (Figure 0.3-b).

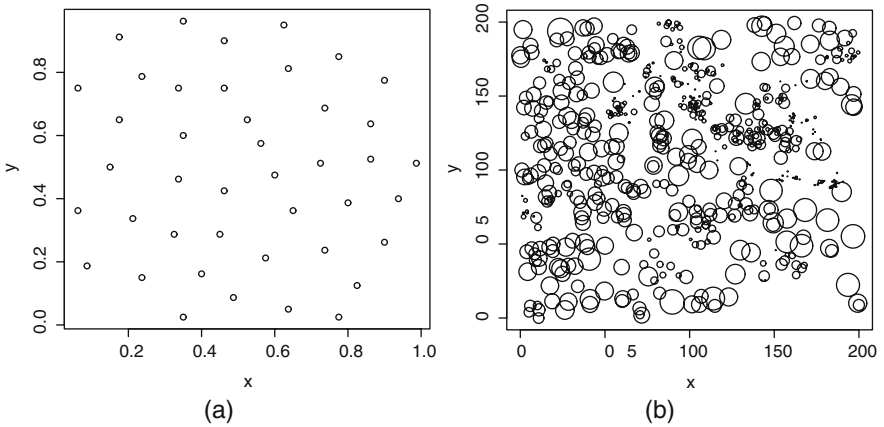

Fig. 0.3 (a) 42 cell centers of a histological section seen under a microscope (`cells` dataset from the `spatstat` package); (b) Position and diameter of 584 pine trees in a forest (`longleaf` dataset from the `spatstat` package).

As is the case for time series, spatial statistics differ from classical statistics due to non-independence of observations; throughout this book, we will generally call X a spatial process or *random field*.

This dependency structure means there is redundancy in available information that can be exploited when making predictions, though it also modifies statistical behavior. Unbiasedness, consistency, efficiency and convergence in distribution of estimators all have to be reexamined in this context. The originality of spatial statistics is to make use of *non-causal modeling*; in this sense, spatial statistics is radically different to time series statistics where causal models use the passage of time and a notion of the "past" (modeling river flows, stock prices, evolution of unemployment rates, etc.). Markov spatial modeling works with the idea of the spatial neighborhood of site s "in all directions." This includes dimension $d = 1$: for example, if $S \subseteq \mathbb{Z}^1$ and X_s is the quantity of corn harvested from each corn stalk along a row, a reasonable model would compare X_s with its *two neighbors*, the stalks to the "left" X_{s-1} and "right" X_{s+1}. We see that causal autoregressive modeling of X_s based on

X_{s-1} has no obvious meaning. If the crop is in a field, we could let the harvested quantity $X_{s,t}$ at site (s,t) depend on that of its 4 nearest neighbors $X_{s-1,t}$, $X_{s+1,t}$, $X_{s,t-1}$ and $X_{s,t+1}$, or even perhaps its 8 nearest neighbors.

These three types of spatial structure (cf. Cressie, (48)) provide the framework to this book. The first three chapters are devoted to modeling each in turn (Chapter 1: Second-order models, geostatistics, intrinsic models and autoregressive models; Chapter 2: Gibbs-Markov random fields over networks; Chapter 3: Spatial point processes). Due to the importance of simulation in spatial statistics, Chapter 4 presents *Monte Carlo Markov Chain* (MCMC) methods for spatial statistics. Chapter 5 then brings together the most important statistical methods for the various models and data types and investigates their properties. Four appendices round things off with a presentation of the most useful probabilistic and statistical tools in spatial statistics (simulation, limit theorems and minimum contrast estimation) as well as software packages for performing analyses presented in the book.

Numerous examples, most of them treated with the *R* software package (178), shed light on the topics being examined. When the data being studied are not directly available in *R* or from some other specified location, descriptions, relevant program scripts and links can be found at the website of the book:

`www.dst.unive.it/~gaetan/ModStatSpat.`

Each chapter ends with a set of exercises.

The bibliography gives the reader the chance to enrich their knowledge of notions only briefly presented here as well as several technical results whose proofs have been omitted. We also list reference books that fill gaps remaining after our intentionally reduced and non-exhaustive treatment of this multi-faceted subject undergoing great development (69).

Our thanks go to all our colleagues who have given us a taste for spatial analysis, for their ideas, remarks, contributions and those who have allowed us to use data collected from their own work. We would equally like to thank the *R Development Core Team* and authors of spatial packages for *R* (178) who have made their powerful and efficient software freely available to the public, indispensable when working with methods and tools described here. We thank reviewers for their careful rereading of the first draft; their remarks have helped to significantly improve the present version. Thanks to Bernard Ycart for encouraging us to expand an initially more modest project. Of course, we could never have undertaken this work without the patience and support of our families and the backing of our respective research teams, Dipartimento di Statistica - Università Ca' Foscari Venezia and Laboratoire SAMOS - Université Paris 1. Lastly, many thanks to Kevin Bleakley for the translation and English adaptation, done with much competence. Any remaining errors are ours.

Venice and Paris, *Carlo Gaetan*
August 2009 *Xavier Guyon*

Contents

Abbreviations and notation

AR	autoregressive
ARMA	autoregressive moving average
a.s.	almost surely
CAR	conditional autoregressive
CFTP	coupling from the past
Ch.	chapter
CLS	conditional least squares
CLT	central limit theorem
c.n.d.	conditionally negative definite
CPL	conditional pseudo-likelihood
CSR	complete spatial randomness
ex.	example
Fig.	figure
GLM	generalized linear model
GLS	generalized least squares
GWN	Gaussian white noise
iff	if and only if
i.i.d.	independent and identically distributed
i.n.i.d.	independent non-identically distributed
LS	least squares
LSE	least squares estimation
MA	moving average
MAP	maximum a posteriori
MCMC	Monte Carlo Markov Chain
MCPL	maximum conditional pseudo-likelihood
MH	Metropolis-Hastings
ML	maximum likelihood
MPM	marginal posterior mode
MPP	marked point process
MSE	mean square error
MSNE	mean square normalized error

NN	nearest neighbor
OLS	ordinary least squares
p.d.	positive definite
PP	point process
PPP	Poisson point process
PRESS	prediction sum of squares
p.s.d.	positive-semidefinite
QGLS	quasi-generalized least squares
q.m.	quadratic mean
resp.	respectively
r.r.v.	real random variable
RSS	residual sum of squares
r.v.	random variable
SA	simulated annealing
SAR	simultaneous autoregressive
SARX	SAR with exogenous variables
SLLN	strong law of large numbers
s.t.	such that
STARMA	spatio-temporal ARMA
SWN	strong white noise
TV	total variation
WLS	weighted least squares
WN	white noise
w.r.t.	with respect to
WWN	weak white noise

$\mathscr{B}(S)$	Borel sets of S, $S \subseteq \mathbb{R}^d$
$\mathscr{B}_b(S)$	bounded Borel sets of S, $S \subseteq \mathbb{R}^d$
$\sharp(A)$	cardinality of (A)
$\lvert\Sigma\rvert$	determinant of Σ
$\delta(A)$	diameter of A : $\delta(A) = \sup_{x,y\in A} d(x,y)$
$\lVert\cdot\rVert$ or $\lVert\cdot\rVert_2$	Euclidean norm on $\mathbb{R}^p$: $\lVert x\rVert = \sqrt{\sum_1^p x_i^2}$
$\langle i,j\rangle$	i and j are neighbors
$[x]$	integer part of x
$d(A)$	interior diameter of A: $d(A) = \sup\{r : \exists x \text{ s.t. } B(x;r) \subseteq A\}$
$\cdots \doteq K(\theta,\alpha)$	$K(\theta,\alpha)$ equals, by definition, the left-hand side
$A \otimes B$	Kronecker product of matrices A and B
$\lVert\cdot\rVert_1$	l^1 norm: $\lVert x\rVert_1 = \sum_1^p \lvert x_i\rvert$
$\lambda_M(B)$	largest eigenvalue of B
ν	Lebesgue measure on $\mathbb{R}^d$
∂A (∂i)	neighborhood border of A (of site i)
$\pi \ll \mu$	π is absolutely continuous with respect to μ
${}^t uv$	scalar product on $\mathbb{R}^p$: ${}^t uv = \sum_{i=1}^p u_i v_i$

$\lambda_m(C)$	smallest eigenvalue of B
$\|\cdot\|_\infty$	sup norm: $\|x\|_\infty = \sup_i \|x_i\|$
$\|\cdot\|_{VT}$	total variation norm
${}^tC$	transpose of C
$X \sim \mathcal{N}(0,1)$	X has a $\mathcal{N}(0,1)$ distribution

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 $S \subset \mathbb{Z}^d$ (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.

C. Gaetan, X. Guyon, *Spatial Statistics and Modeling,* Springer Series in Statistics,
DOI 10.1007/978-0-387-92257-7_1,

1.1 Some background in stochastic processes

Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space, S a set of sites and $(E, \mathscr{E})$ 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 $(\Omega, \mathscr{F}, \mathbb{P})$ and taking values in $(E, \mathscr{E})$. $(E, \mathscr{E})$ 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 $\omega \longmapsto (X_{s_1}(\omega), X_{s_2}(\omega), \ldots, X_{s_n}(\omega))$: that is, for $A_i \in \mathscr{E}$, $i = 1, \ldots, n$,

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

The event $(X_{s_1} \in A_1, X_{s_2} \in A_2, \ldots, X_{s_n} \in A_n)$ of $\mathscr{E}$ 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 $\mathscr{F}$. 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-σ-algebra $\mathscr{A} \subseteq \mathscr{F}$ generated by the set of cylinders of $\mathscr{E}$ (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 σ-field $\mathscr{E} = \mathscr{B}(E)$.

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

With $L^2 = L^2(\Omega, \mathscr{F}, \mathbb{P})$ representing the set of real-valued and square integrable random variables on $(\Omega, \mathscr{F})$, $X \in L^2$ 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:

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

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.

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 $m_\Lambda = E(X_\Lambda)$ is the mean of $X_\Lambda = (X_s, s \in \Lambda)$ and Σ_Λ its covariance, then if Σ_Λ is invertible, the density (or likelihood) of X_Λ with respect to the Lebesgue measure on $\mathbb{R}^{\sharp\Lambda}$ is

$$f_\Lambda(x_\Lambda) = (2\pi)^{-\sharp\Lambda/2}(\det\Sigma_\Lambda)^{-1/2}\exp\left\{-\frac{1}{2}{}^t(x_\Lambda - m_\Lambda)\Sigma_\Lambda^{-1}(x_\Lambda - m_\Lambda)\right\},$$

where $\sharp U$ is the cardinality of U and x_Λ possible values of X_Λ. 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 $\mathbb{R}^2$, $S = \{ne_1 + me_2, n \text{ and } m \in \mathbb{Z}\}$ 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:

$$\forall s,t \in S\text{: } E(X_s) = m \text{ and } c(s,t) = Cov(X_s, X_t) = C(t-s).$$

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

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

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. $\forall h \in S$, $|C(h)| \leq C(0) = Var(X_s)$.
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, ... are stationary covariances, the following functions are as well:*
   a. $C(h) = a_1 C_1(h) + a_2 C_2(h)$ *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 μ 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_μ 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:

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

(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 $X_t = \sqrt{a_1}X_{1,t} + \sqrt{a_2}X_{2,t}$ (resp. $X_t = X_{1t}X_{2t}$). □

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 $s \neq t$, $Cov(X_s, X_t) = 0$ and $Var(X_s) = \sigma^2 < \infty$. 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 $s,t \in S$, $Cov(X_s, X_t)$ depends only on $\|s-t\|$:

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

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 $\|h\|$,

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

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

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

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 $\mathscr{B}(\mathbb{R}^d)$ such that:

$$C(h) = \int_{\mathbb{R}^d} e^{i^t hu} F(du), \tag{1.2}$$

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

$$f(u) = (2\pi)^{-d} \int_{\mathbb{R}^d} e^{-i\,{}^t hu} C(h)dh.$$

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:

$$\begin{aligned} C(h) = c_d(r) &= \int_{[0,\infty[} \left[\int_{S_d} \cos(r\rho\, {}^t\theta\alpha) d\sigma(\alpha) \right] \rho^{d-1} f_d(\rho) d\rho \\ &= \int_{[0,\infty[} \Lambda_d(r\rho) \rho^{d-1} f_d(\rho) d\rho. \end{aligned} \tag{1.3}$$

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), $\Lambda_d(v) = \Gamma(d/2)(v/2)^{-(d-2)/2} \mathscr{J}_{(d-2)/2}(v)$, where $\mathscr{J}_\kappa$ is the Bessel function of the first kind of order κ (2). For $n = 1$, 2 and 3, we have:

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

Using (1.3), we obtain lower bounds:

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

In particular, we get the lower bounds (227; 184), tighter than those in (1.1): $\rho_0(\|h\|) \geq -0.403$ in $\mathbb{R}^2$, $\rho_0(\|h\|) \geq -0.218$ in $\mathbb{R}^3$, $\rho_0(\|h\|) \geq -0.113$ in $\mathbb{R}^4$ and $\rho_0(\|h\|) \geq 0$ in $\mathbb{R}^{\mathbb{N}}$.

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:

$$f(u) = \frac{1}{2\pi} \int_{]-\infty,\infty[} be^{-\alpha|t|-iut} dt = \frac{\alpha b}{\pi(\alpha^2+u^2)}.$$

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

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

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 $\mathbb{T}^d = [0, 2\pi[^d$ 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 $\mathscr{B}(\mathbb{T}^d)$ such that:

$$C(h) = \int_{\mathbb{T}^d} e^{i^t uh} F(du).$$

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 $L^2(\mathbb{T}^d)$:

$$f(u) = (2\pi)^{-d} \sum_{h\in\mathbb{Z}^d} C(h) e^{-i^t uh}. \tag{1.4}$$

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 $f \in \mathscr{C}^k(\mathbb{T}^d)$ where $k = (k_1, \ldots, k_d) \in \mathbb{N}^d$,

$$\limsup_{h\longrightarrow\infty} h^k |C(h)| < \infty,$$

where $h = (h_1, h_2, \ldots, h_d) \longrightarrow \infty$ means at least one coordinate $h_i \to \infty$ and $h^k = h_1^{k_1} \times \ldots \times h_d^{k_d}$. 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 ($X_s = a + bs + \varepsilon_s$ for a stationary centered residual process ε). 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 $\Delta X^{(h)} = \{\Delta X_s^{(h)} = X_{s+h} - X_s : s \in S\}$ is second-order stationary. The semi-variogram of X is the function $\gamma : S \to \mathbb{R}$ defined by:

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

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 $\Delta^k X^{(h)}$ 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 $Var(X_{(u,v)+(1,1)} - X_{(u,v)}) = u + v + 1$ 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 γ 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 γ is continuous at 0, then γ is continuous at every site s where γ is locally bounded.*
6. *If γ 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 $\sum_{i=1}^n a_i X_{s_i} = \sum_{i=1}^n a_i Y_{s_i}$ and

$$Var\left(\sum_{i=1}^n a_i X_{s_i}\right) = \sum_{i=1}^n \sum_{j=1}^n a_i a_j C_Y(s_i, s_j) = -\sum_{i=1}^n \sum_{j=1}^n a_i a_j \gamma(s_i - s_j) \geq 0.$$

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

$$2\gamma_Y(h) = Var(X_{A(s+h)} - X_{As}) = 2\gamma(Ah).$$

(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 each of 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, γ is uniformly continuous on any set over which γ is bounded. (6) We prove by induction that for each $n \in \mathbb{N}$ and $h \in \mathbb{R}^d$, $\gamma(nh) \leq n^2\gamma(h)$. 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 $\widetilde{x} = \delta \|x\|^{-1}$, the decomposition $x = n\widetilde{x} + \tau$ defines some τ satisfying $\|\tau\| \leq \delta$. We conclude by remarking that

$$\begin{aligned}\gamma(x) &= \gamma(n\widetilde{x} + \tau) \leq \gamma(n\widetilde{x}) + \gamma(\tau) + 2\sqrt{\gamma(n\widetilde{x})\gamma(\tau)} \\ &\leq Cn^2 + C + 2Cn = C(n+1)^2 \leq C\left(\frac{\|x\|}{\delta} + 1\right)^2. \qquad \square\end{aligned}$$

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 $Var(Z_1) = \sigma_1^2 > 0$.

Characterizations exist to ensure a function γ is a variogram, one of them being the following (43): if γ is continuous and if $\gamma(0) = 0$, then γ 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

$$2\gamma(h) = 2(C(0) - C(h)). \tag{1.5}$$

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 $C(0) = Var(X)$ 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.

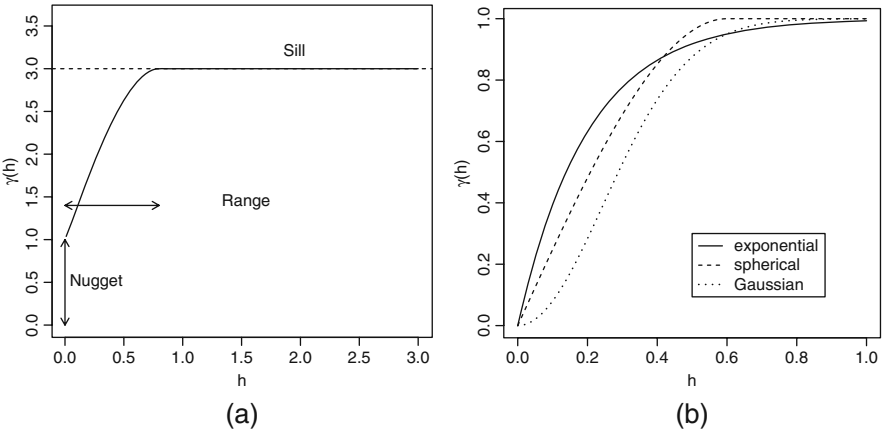

Fig. 1.1 (a) Semivariogram of a stationary model with a nugget effect component; (b) variogram models that have the same range.

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 σ^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* ($d \leq 3$):

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

-*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*:

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

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 ε is a white noise in L^2 (nugget effect at the origin) uncorrelated with X whose variogram is continuous and with sill

$$2\gamma_Y(h) = 2\sigma_\varepsilon^2(1 - \delta_0(h)) + 2\gamma_X(h).$$

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:

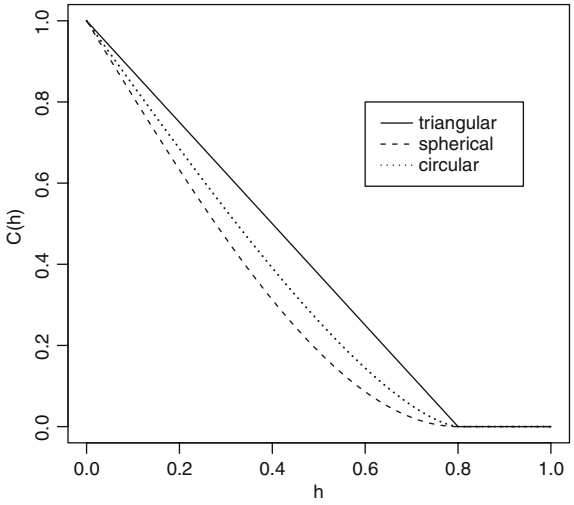

Fig. 1.2 Graph showing triangular, spherical and circular covariances with $\sigma^2 = 1$ and $a = 0.8$.

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

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 $\mathbb{R}^2$:

$$C_{circ}(h;a,\sigma^2) = \begin{cases} \frac{2\sigma^2}{\pi}\left(\arccos\ \frac{\|h\|}{a} - \frac{\|h\|}{a}\sqrt{1-\left(\frac{\|h\|}{a}\right)^2}\right) & \text{if } \|h\| \le a \\ 0 & \text{otherwise} \end{cases}. \quad (1.6)$$

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 $\mathbb{R}^1$:

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

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 $\mathbb{R}^{d'}$, $d' \le 3$, 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.

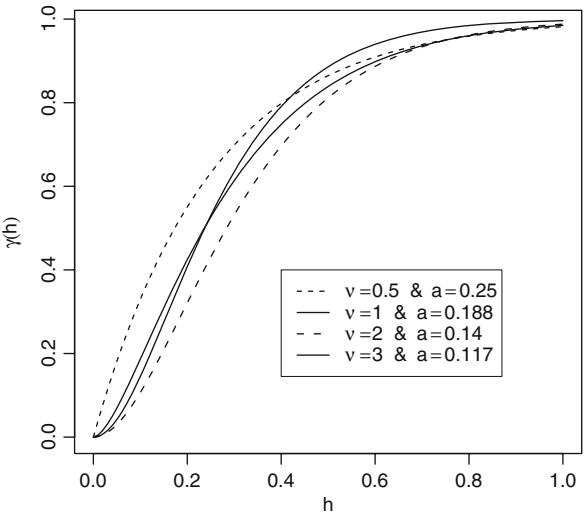

Fig. 1.3 Matérn semivariograms with the same range but different ν.

4. Our interest in Matérn covariance is due to its parameter ν 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 $\widehat{X}$ using kriging (cf. §1.9): increasing ν increases regularity of γ at 0 and regularity of the process X (the kriging surface $\widehat{X}$). 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 α 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):

$$2\gamma(h) = \sum_{j=1}^{K} 2\gamma_j(h) \qquad (\text{resp. } C(h) = \sum_{j=1}^{K} C_j(h)).$$

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

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

We say that a variogram is *anisotropic* if at least two directional variograms 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 2γ 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 $Q = {}^t AA$. 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 $(\lambda_k, k = 1,\ldots,d)$, $\gamma(\widetilde{h}) = \gamma_0(\sum_{k=1}^{d} \lambda_k \widetilde{h}_k)$ in these new coordinates $\widetilde{h}$.

For example, if A is a rotation of angle ϕ 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 γ_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

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

where g is a bijective (or anamorphic) deformation of the space S (cf. (170; 171) for examples of such deformations).

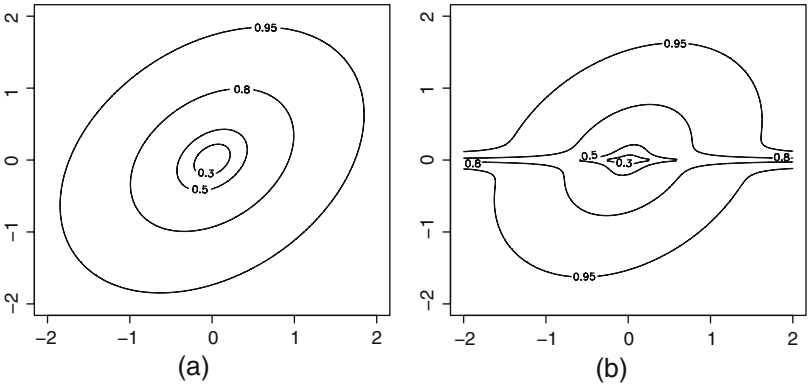

Fig. 1.4 (a) Geometric anisotropy and (b) zonal (or stratified) anisotropy.

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 γ will thus be direction-dependent (cf. Fig. 1.4-b). We say we have *zonal anisotropy* or *stratified anisotropy* if γ 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 σ_1^2 in the $(1,0)$ direction, where σ_i^2 are the sills of γ_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 $X_s = X_x^1 + X_y^2$, with $Cov(X_x^1, X_y^2) = 0$ and $s = {}^t(x,y)$, then $\gamma(h) = \gamma_1(h_1) + \gamma_1(h_2)$ and for $h_x = {}^t(d_x,0)$, $h_y = {}^t(0,d_y)$, $X_s - X_{s+h_x} - X_{s+h_y} + X_{s+h_x+h_y} \equiv 0$.

More generally, anisotropy can be obtained by combining other anisotropies. Figure 1.4-b gives an example where γ_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 γ_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 $s_n \longrightarrow s$ in S, $E(X_{s_n} - X_s)^2 \to 0$.

The following proposition characterizes q.m. continuity.

Proposition 1.3. *Let X be a centered L^2 process with covariance $C(s,t) = Cov(X_s, X_t)$. 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 $E(X_{s_0+h} - X_{s_0})^2 \to 0$ as $h \to 0$. In effect:

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

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 $e_3 \to 0$ if h and $k \to 0$ and C is continuous on the diagonal. □

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

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

When X is an intrinsic Gaussian process, this continuity holds if $\gamma(h) \leq c\,|\log \|h\||^{-(1+\varepsilon)}$ 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 $S \subset \mathbb{R}^1$ 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 $C(s,t) = Cov(X_s, X_t)$. If $\dfrac{\partial^2}{\partial s \partial t} C(s,t)$ exists and is finite on the diagonal*