Some aspects of geostatistics

Rasmus Waagepetersen

Contents

Introduction

Geostatistics is a branch of statistics which include a wide range of methods for simulation and prediction of spatial stochastic processes. The aim of these notes is to give a survey of a selection of topics related geostatistics. Many important issues like e.g. estimation are not or only very cursorily treated, and the selection of topics considered here, is largely determined by those topics which I went through at the Ph.D.-course on Geostatistics in the autumn 1993 at the Dept. of Theoretical Statistics. More thorough references on geostatistics are Journel & Huijbregts (1978) and Cressie (1991). The last and newest reference contains a vast amount of informations on the subject, but could have been more well-arranged.

Chapter 1 in these notes deals with linear prediction of second order stationary and intrinsically stationary stochastic processes. In connection with linear prediction of such processes the covariogram, semivariogram or the generalized covariance must be modelled and it is therefore important to be able to decide for a given candidate function whether this function is a valid covariogram, smivariogram or generalized covariance model. Chapter 2 is concerned with this problem. Finally, in Chapter 3, various methods for simulation of Gaussian random fields are considered. Simulation is an important tool for model selection and Monte Carlo calculations of analytically intractable problems.

1 Optimal linear prediction of nonstationary processes

1.1 Some basic definitions

In these notes three types of stochastic processes will be considered. We let

$$Z = \{Z(s) : s \in D\}$$

denote a stochastic process with $D \subseteq \mathbf{R}^d$ and $Z(s) \in \mathbf{R}$. Under the assumption $\forall\, s \in D : E(Z(s))^2 < \infty$, then Z is *second order stationary* if

$$\forall\, s \in D : E(Z(s)) = \mu \tag{1.1.1}$$

$$\forall\, s,t \in D : Cov(Z(s), Z(t)) = C(s-t) \tag{1.1.2}$$

$C(\cdot)$ is called the *covariogram* of Z. If Z is Gaussian, that is, all the finitely dimensional distributions are Gaussian, then Z is fully characterized by the mean and covariance, and in this case (1.1.1) and (1.1.2) also imply strict stationarity (this means, that the finitely dimensional distributions are translation invariant).

If we only assume $\forall\, s,t \in D : E(Z(s) - Z(t))^2 < \infty$ and replace (1.1.2) by

$$\forall\, s,t \in D : Var(Z(s) - Z(t)) = 2\gamma(s-t) \tag{1.1.3}$$

then Z is *intrinsically stationary*, and $2\gamma(\cdot)$ is called the *variogram* of Z. The assumption of intrinsically stationarity is less restrictive than that of second order stationarity, and the class of linear predictors available for intrinsically stationary processes is not essentially smaller than the class of linear predictors for second order stationary processes. This hypothesis is therefore often preferred in practice.

An intrinsically stationary stochastic process is a special case of a class of stochastic processes called the *intrinsic random functions* of order k or IRF-k where $k \in \mathbf{N}_0$. The stochastic process Z is an IRF-k, if the stochastic process Y given by

$$Y(s) = \sum_{i=1}^{n} \lambda_i Z(s_i + s)$$

is second order stationary for all $n \in \mathbf{N}$, $s_1 = (s_{11}, \ldots, s_{1d}), \ldots, s_n = (s_{n1}, \ldots, s_{nd}) \in D$ and all $\lambda = (\lambda_1, ..., \lambda_n) \in \mathbf{R}^n$ such that

$$\sum_{i=1}^{n} \lambda_i s_{i1}^{a_1} \ldots s_{id}^{a_d} = 0$$

for all $a = (a_1, \ldots, a_d) \in \mathbf{N}_0^d : a. \sum_{j=1}^{d} a_j \leq k$. The latter condition could also be put in matrix form

$$\lambda X = 0 \tag{1.1.4}$$

with

$$X = \left(s_{i1}^{a_1} \cdots s_{id}^{a_d}\right)_{\substack{i=1,\ldots,n, \\ (a_j)_{1\leq j\leq d} \in \mathbf{N}_0^d : a. \leq k}} = \begin{bmatrix} 1 & s_{11} & - & s_{11}^{a_1} \cdots s_{1d}^{a_d} & - & s_{1d}^{k} \\ | & & & & & | \\ 1 & s_{n1} & - & s_{n1}^{a_1} \cdots s_{nd}^{a_d} & - & s_{nd}^{k} \end{bmatrix} \tag{1.1.5}$$

For an IRF-k there exists a function $K(\cdot)$ called the generalized covariance (Matheron (1973)) such that for every λ where (1.1.4) holds, the variance of $\sum_{i=1}^{n} \lambda_i Z(s_i)$ is given by

$$Var\left(\sum_{i=1}^{n} \lambda_i Z(s_i)\right) = \sum_{i=1}^{n}\sum_{j=1}^{n} \lambda_i \lambda_j K(s_i - s_j)$$

An intrinsically stationary process Z is an IRF-0 as easy calculations show: AConsider $Y(s) = \sum_{i=1}^{n} \lambda_i Z(s_i + s)$ where $\sum_{i=1}^{n} \lambda_i = 0$. The mean of $Y(s)$ is then 0 and the covariance of $Y(s)$ is (with s, t and $s_0 \in D$)

$$Cov(Y(s), Y(t)) = Cov\left(\sum_{i=1}^{n} \lambda_i (Z(s_i + s) - Z(s_0)), \sum_{j=1}^{n} \lambda_j (Z(s_j + t) - Z(s_0))\right) =$$

$$\sum_{i=1}^{n}\sum_{j=1}^{n} \lambda_i \lambda_j (\gamma(s_i + s - s_0) + \gamma(s_j + t - s_0) - \gamma(s_i - s_j + (s - t))) =$$

$$-\sum_{i=1}^{n}\sum_{j=1}^{n} \lambda_i \lambda_j \gamma((s_i - s_j) + (s - t))$$

Thus, Y is second order stationary and it follows, that the intrinsically stationary process Z is an IRF-0. An IRF-0 need on the other hand not to be intrinsically stationary. For an IRF-0 Z let for $s, t \in D$, $Y(u) = Z(s + u) - Z(t + u)$. Then for all $s', t' \in D$ with $s - t = s' - t'$

$$Var(Y(0)) = Var(Y(s' - s)) \Leftrightarrow Var(Z(s) - Z(t)) = Var(Z(s') - Z(t'))$$

so that (1.1.3) holds. The mean of $Z(s)$, $s \in D$ may however not be finite, and in the case where it is finite, it could be nonconstant. In the case $D = \mathbf{R}^1$, $E(Z(s))$, $s \in D$ could e.g. be a linear function of s. Note, that for $k \geq 0$, the class of IRF-k's is contained in the class of IRF-$k + 1$'s. That is the assumption that a random function is an IRF-k becomes weaker as k increases.

1.2 Linear prediction of nonstationary processes

Assume that $Z = \{Z(s)|s \in D\}$ is a stochastic process with a nonstationary mean. Given observations $Z(s_1), \ldots, Z(s_n)$ of Z at locations $s_1, \ldots s_n \in D$, one may wish to find a predictor of $Z(s_0)$ at some location s_0 where Z is not observed. An *optimal linear predictor* $\hat{Z}(s_0)$ is a linear combination (or weighted average) $\sum_{i=1}^{n} \lambda_i Z(s_i)$ of $Z(s_1), \ldots, Z(s_n)$, such that $\hat{Z}(s_0)$ is

(1.2.1) unbiased, i.e. $E(\hat{Z}(s_0)) = E(Z(s_0))$ and

(1.2.2) among all linear predictors, $\hat{Z}(s_0)$ minimizes the *mean square prediction error* (m.s.p.e.), i.e. $E(\hat{Z}(s_0) - Z(s_0))^2 \leq E(Z(s_0) - \sum_{i=1}^{n} \alpha_i Z(s_i))^2)$ for every linear combination $\sum_{i=1}^{n} \alpha_i Z(s_i)$.

The predictor that minimizes the m.s.p.e. among all predictors of $Z(s_0)$ is the conditional expectation $E(Z(s_0)|Z(s_1), \ldots, Z(s_n))$. For a zero-mean Gaussian process, the conditional expection is a linear combination of $Z(s_1), \ldots, Z(s_n$, and the optimal linear predictor is therefore in this case also the predictor which minimizes the m.s.pe. among all predictors.

One way to model Z is by

$$Z(s) = \beta f(s)^t + \delta(s), \tag{1.2.3}$$

where $\beta = (\beta_1, \ldots, \beta_q)$ is a vector of regression coefficients, $f(s) = (f_1(s), \ldots, f_q(s))$ is a known function from $\mathbf{R}^d$ into $\mathbf{R}^q$ and $\delta(\cdot)$ is a zero-mean stochastic process where either the covariances $C(s,t) = Cov(Z(s), Z(t))$, $s,t \in D$ or the variances of increments $2\gamma(s,t) = Var(Z(s) - Z(t))$, $s,t \in D$ exist. If we assume that the covariance exists and is known, the optimal linear predictor can be expressed in terms of the covariance function $C(\cdot,\cdot)$ as

$$\hat{Z}(s_0) = \hat{\beta} f(s_0)^t + \Big(Z_{dat} - \hat{\beta}F\Big)C^{-1}c^t \tag{1.2.4}$$

where $Z_{dat} = (Z(s_1), \ldots, Z(s_n))$, $C = (C(s_i, s_j))_{i,j=1,\ldots,n}$, $c = (C(s_0, s_i))_{i=1,\ldots,n}$, $F = (f^t(s_1), \ldots, f^t(s_n))$ and

$$\hat{\beta} = Z_{dat}C^{-1}F^t\big(FC^{-1}F^t\big)^{-1} \tag{1.2.5}$$

This is of course only valid when C is positive definite and F has full rank. In the case where $C(s,s)$, $s \in D$, is constant, the predictor (1.2.4) can also be written in terms of the variances of increments $2\gamma(s,t) = Var(Z(s) - Z(t)$, $s,t \in D$, by using the relation

$$2\gamma(s,t) = C(s,s) + C(t,t) - 2C(s,t).$$

The predictor (1.2.4) is obtained by minimizing the m.s.p.e. under the unbiasedness condition (1.2.1) using Lagrange multipliers, see e.g. Cressie (1991, p. 154). . The

unbiasedness condition here takes the form $(\lambda_1, \ldots, \lambda_n)F^t = f(s_0)$. Another way to arrive at the predictor (1.2.4) is to consider $(Z(s_0), Z_{dat})$ as Gaussian distributed

$$(Z(s_0), Z_{dat}) \sim N\Big(\beta\Big(f(s_0)^t, F\Big), \begin{pmatrix} C(s_0, s_0) & c \\ c^t & C \end{pmatrix}\Big).$$

Then (1.2.5) is the maximum likelihood estimate of β, and by substitution of β with $\hat{\beta}$ in the conditional expectation of $Z(s_0)$ given Z_{dat}, the predictor (1.2.4) is obtained. This approach did of course not á priori guarantee to lead to the optimal linear predictor.

If only the variances of increments exist we must demand that $\sum_{i=1}^n \lambda_i = 1$ in order that $E(Z(s_0) - \sum_{i=1}^n \lambda_i Z(s_i))^2$ is well defined. Minimizing the mean square prediction error under this constraint on the weights λ_i and the unbiasedness condition (1.2.1) gives

$$\hat{Z}(s_0) = \hat{\beta} f(s_0)^t + \Big(Z_{dat} - \hat{\beta}F\Big)\Gamma^{-1}\gamma^t \tag{1.2.6}$$

where F is as before, $\Gamma = (\gamma(s_i, s_j))_{i,j=1,\ldots,n}$, $\gamma = (\gamma(s_0, s_i))_{i=1,\ldots,n}$ and

$$\hat{\beta} = Z_{dat}\Gamma^{-1}F^t(F\Gamma^{-1}F^t)^{-1}$$

Also in this case, Lagrange multipliers can be used for the minimization, c.f. Cressie (1991, p. 152). Note that the constraint $\sum_{i=1}^n \lambda_i = 1$ is implied by the unbiasedness condition whenever one of the functions $f_i(\cdot)$, $i = 1, \ldots, q$, is identically one. Normally this is not a severe restriction.

Another way to model nonstationarity is by modelling Z as an IRF-k as defined in the first section. In this case only linear predictors $\sum_{i=1}^n \lambda_i Z(s_i)$ where $\lambda = (-1, \lambda_i, \ldots, \lambda_n)$ satisfies (1.1.4) of Section 1 are available. The m.s.pe. may otherwise not be well-defined. A linear predictor is thus found by minimizing the m.s.p.e under the constraints

$$\sum_{i=1}^n \lambda_i s_{i1}^{a_1} \cdots s_{id}^{a_d} = s_{01}^{a_1} \cdots s_{0d}^{a_d} \tag{1.2.7}$$

for every $(a_1, \ldots, a_d) \in \mathbf{N}^d : \sum_{i=1}^n a_i \leq k$. Under these constraints the mean square prediction error is well-defined and given in terms of the generalized covariance $K(\cdot)$. Note, that (1.2.7) does not guarantee unbiasedness. Assuming that $K(\cdot)$ is known, that X is of maximal rank, and that K is invertible, the linear predictor is given by

$$\hat{Z}(s_0) = \hat{B}x^t + \Big(Z_{dat} - \hat{B}X\Big)K^{-1}k^t \tag{1.2.8}$$

(c.f. Delfiner, 1976), where $x = (s_{01}^{a_1} \cdots s_{0d}^{a_d})_{(a_1,\ldots,a_d)\in\mathbf{N}_0^d : a_\cdot \leq k} = (s_{01}, \ldots, s_{01}^{a_1} \cdots s_{0d}^{a_d}, \ldots, s_{0d}^k)$, $K = (K(s_i, s_j)_{i,j=1,\ldots,n})$, $k = (K(s_0, s_i)_{i=1,\ldots,n})$, X is given by (1.1.5) and

$$\hat{B} = Z_{dat}K^{-1}X(X^tK^{-1}X)^{-1}.$$

If Z is of the form

$$Z(s) = \beta f(s)^t + \delta(s)$$

where $f(s)$ is a vector of mononomials of degree less than k in the coordinates of s (i.e. $f_j(s) = s_1^{a_{1j}} \cdots s_d^{a_{dj}}$, $\sum_{i=1}^{d} a_{ij} \leq k$, $j = 1, \ldots q$) and $\delta(\cdot)$ is a IRF-k where the mean exists and is zero, then the constraints (1.2.7) entails unbiasedness.

Prediction using the predictors (1.2.4) and (1.2.6) is termed universal kriging, whereas prediction using (1.2.8) is called IRF-kriging.

1.3 Comparison of universal kriging and IRF kriging

If we consider an IRF-k Z where Z can be represented as

$$Z(s) = \mu(s) + \delta(s) \tag{1.3.1}$$

where $\mu(s)$ is a deterministic trend and $\delta(s)$ is a zero-mean second order process then Z posses both a generalized covariance $K(\cdot)$ and an ordinary covariance function $C(\cdot, \cdot)$. These to functions are related by

$$Var\left(\sum_{i=1}^{n} \lambda_i Z(s_i)\right) = \sum_{i=1}^{n}\sum_{j=1}^{n} \lambda_i \lambda_j C(s_i, s_j) = \sum_{i=1}^{n}\sum_{j=1}^{n} \lambda_i \lambda_j K(s_i - s_j)$$

for all $\lambda = (\lambda_1, \ldots, \lambda_n)$ satisfying (1.1.4). In other words, with X again given by (1.1.5), $L = \text{span(rows of } X)$, and P the projection on $L^{\perp}$, then $C = (C(s_i, s_j))_{i,j=1,\ldots,n}$ and $K = (K(s_i, s_j))_{i,j=1,\ldots,n}$ are related by

$$PCP = PKP$$

Christensen (1990) considers the model (1.3.1) where $\mu(s)$ is given by $\beta f(s)^t$ where $f(s)$ is a polynomial in the coordinates of s. He shows that the predictors obtained by universal kriging and IRF-kriging are equivalent. Christensen (1990) argues that the crucial point concerning the hypothesis that Z is an IRF is that it allows for a nonstationary covariance, while it is not so important that it allows for nonexistence of the covariance. His argument follows these lines: observable random variables are bounded because our ability to observe them is bounded. The variances of an observable process therefore exist. This argument does not take into account the possibility of considering some observations as censored by the observation method. Christensen (1990) does not assume that the trend model includes all mononomials of order less or equal than k. This is a mistake because then the unbiasedness condition $\lambda X = f(s_0)$ does not imply the conditions (1.2.7) which are necessary in order that $Var\left(Z(s_0) - \sum_{i=1}^{n} \lambda_i Z(s_i)\right)$ is given in terms of the generalized covariance K, so that the IRF-predictor can be obtained.

Sofar we have assumed that either $C(\cdot, \cdot)$, $\gamma(\cdot, \cdot)$ or $K(\cdot)$ is known. In practice, when applying universal kriging or IRF-kriging, these functions are unknown and must be replaced by suitable estimates. In order to be able to estimate $C(\cdot, \cdot)$ or $\gamma(\cdot, \cdot)$ from a single outcome of Z it is in most cases necessary to make the simplifying assumptions that the fluctuation process $\delta(\cdot)$ is either second order stationary or intrinsically

stationary. Recalling the result of Christensen (1990) it seems, that the real difference between universal kriging and IRF-kriging appears when the covariance, variance of increments or the generalized covariance of the fluctuation process $\delta(\cdot)$ is to be modelled. Therefore IRF-kriging, where it is only necessary to assume second order stationarity of certain generalized increments may seem more generally applicable than universal kriging. On the other hand universal kriging offers a larger class of trend models than the polynomial models which one has in mind when applying IRF-kriging.

1.4 Estimation of the fluctation process

Assume that the fluctuation process δ from Section 1.3 is either second order or intrinsically stationary. One particular problem concerning the estimation of the covariogram $C(\cdot)$ or variogram $2\gamma(\cdot)$ are biased estimators, where the bias is caused by the presence of the nonstationary mean. Armstrong (1984) remarks, that this is the case when one applies the empirical estimators of the variogram or the covariogram given by

$$2\hat{\gamma}(h) = \frac{1}{N(h)} \sum_{(i,j)\in N(h)} \left(Z(s_i) - \hat{\beta} f(s_i)^t - Z(s_j) - \hat{\beta} f(s_j)^t \right)^2$$

and

$$\hat{C}(h) = \frac{1}{N(h)} \sum_{(i,j)\in N(h)} \left(Z(s_i) - \hat{\beta} f(s_i)^t \right) \left(Z(s_j) - \hat{\beta} f(s_j)^t \right)$$

where $N(h) = \{(i,j) : \| s_i - s_j \| = h\}$ and $\hat{\beta}$ is some initial estimate of β. Delfiner (1976) also mentions this problem. Unfortunately none of these two references mentions how the initial estimator $\hat{\beta}$ is obtained (if a bad $\hat{\beta}$ is chosen, it does not seem surprising that the empirical estimators will be biased). If a parametric model for the covariogram has been chosen, i.e.

$$C(\cdot) = C(\cdot; \theta)$$

where θ is some unknown parameter, maximum likelihood estimation under the assumption of Gaussianity could be tried, but also in this case the resulting estimates tends to be biased (Cressie, 1991, p. 92)

A way to deal with the biasedness problems is to base inference on contrasts λZ_{dat}^t. Here λ is a vector such that

$$E\left(\lambda Z_{dat}^t\right) = 0$$

that is $\lambda \in$ span(rows of $F)^{\perp}$ with F as in Section 1.2. Under the assumption that Z is Gaussian such that

$$Z_{dat} \sim N(\beta F, C(\theta))$$

where $C(\theta) = (C(s_i, s_j; \theta))_{i,j=1,\ldots,n}$, then restricted maximum likelihood estimation (REML) is performed by maximum likelihood estimation of θ based on the transformed

data

$$Y_{dat} = \left(\lambda_1 Z^t_{dat}, \ldots, \lambda_{n-q} Z^t_{dat}\right)$$

where $\lambda_1 Z^t_{dat}, \ldots, \lambda_{n-q} Z^t_{dat}$ are linearly independent error contrasts, that is $\lambda_1, \ldots, \lambda_{n-q}$ are linearly independent vectors in $\mathbf{R}^n$, and $q = \text{rank}(F)$. If we denote by Λ the matrix with rows $\lambda_1, \ldots, \lambda_{n-q}$ then the distribution of Y is Gaussian

$$Y_{dat} \sim N\left(0, \Lambda C(\theta) \Lambda^t\right)$$

and does not depend on β, and we can hope to obtain less biased results from the maximum likelihood estimation based on the data Y_{dat}. The idea of REML bears resemblance to that of IRF kriging in that we in both cases consider linear combinations of the data where a nonstationary mean is filtered out.

Another estimation method based on contrasts is minimum norm quadratic (MINQ) estimation as described in Stein (1987). This method demands that the semivariogram is of the form

$$\gamma(\cdot) = \sum_{i=1}^{k} \theta_i \gamma_i(\cdot)$$

where the $\gamma_i's$ are semivariograms and $\theta = (\theta_1, \ldots, \theta_k)$.

Typically when applying IRF-kriging polynomial models are used for the generalized covariance and in this case the coefficients of the polynomium can be determined by minimizing an weighted sum of squares corresponding to a set of regression equations, see Delfiner (1976). Alternatively one can assume that the generalized increments are Gaussian and then perform maximum likelihood estimation.

2 Validity of models for the variogram and the covariogram

In most practical applications of models based on the stochastic processes defined in the first chapter we have to model the covariogram $C(\cdot)$, the variogram $2\gamma(\cdot)$ or the generalized covariance $K(\cdot)$ by some parametric function. In this chapter different criteria for a given function to be a valid covariogram or semivariogram will be considered.

2.1 Valid covariograms and semivariograms

First of all it is of course necessary to define what is meant by a valid covariogram or variogram model. The following definitions are reasonable as will be shown subsequently.

V1 $C : \mathbf{R}^d \to \mathbf{R}$ is a valid covariogram if $C(\cdot)$ is symmetric and nonnegative definite.
V2 $\gamma : \mathbf{R}^d \to \mathbf{R}$ is a valid semivariogram if $\gamma(0) = 0$, γ is symmetric and $-\gamma(\cdot)$ is conditionally nonnegative definite.

That $C(\cdot)$ is nonnegative definite means that for every $n \in \mathbf{N}$, $s_1, \ldots, s_n \in \mathbf{R}^d$, and $(a_1, \ldots, a_n) \in \mathbf{R}^n$,

$$\sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j C(s_i - s_j) \geq 0 \tag{2.1.1}$$

and $-\gamma(\cdot)$ is conditionally nonnegative definite if for every $n \in \mathbf{N}$, $s_1, \ldots, s_n \in \mathbf{R}^d$, and $(a_1, \ldots, a_n) \in \mathbf{R}^n$ with $\sum_{i=1}^{n} a_i = 0$,

$$-\sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j \gamma(s_i - s_j) \geq 0 \tag{2.1.2}$$

The definitions V1 and V2 entails nonnegative prediction variances when a valid covariogram or variogram model is used for linear prediction as described in Chapter 1.

The definitions V1 and V2 will now be justified. We consider first V1. V1 is a natural definition of validity because the following two conditions are equivalent:

a) $C(\cdot)$ is the covariogram of some second order stationary process $Z = \{Z(s) : s \in D\}$.
b) $C(\cdot)$ is symmetric and nonnegative definite.

That a) implies b) follows because given $n \in \mathbf{N}$, $s_1, \ldots, s_n \in \mathbf{R}^d$, and $(a_1, \ldots, a_n) \in \mathbf{R}^n$, we have

$$0 \leq Var\left(\sum_{i=1}^{n} a_i Z(s_i)\right) = \sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j C(s_i - s_j)$$

and it is clear that $C(\cdot)$ is symmetric. To show that b) implies a) we simply construct a Gaussian process with $C(\cdot)$ as its covariogram. For every $s_1, \ldots, s_n \in \mathbf{R}^d$ we let $\Sigma_{s_1,\ldots,s_n} = (C(s_i - s_j))_{i,j=1,\ldots,n}$, and $\Sigma_{s_1,\ldots,s_n}$ is then a nonnegative definite and symmetric matrix to which there corresponds a n-dimensional Gaussian distribution $N(0, \Sigma_{s_1,\ldots,s_n})$. The set of distributions obtained by varying n and $s_1, \ldots, s_n$ is a consistent set of distributions and by Kolmogorovs consistency theorem there exists a stochastic process Z on $\mathbf{R}^d$ with these Gaussian distributions as the finite-dimensional distributions. Z is second order stationary, and the covariogram of Z is of course $C(\cdot)$.

If we assume that $C(\cdot)$ is also continuous we can also show that b) implies a) in another way. Bochners theorem (see the Appendix) states that for $C(\cdot)$ symmetric and continuous, nonnegative definiteness of $C(\cdot)$ is equivalent to:

$$\forall\, t \in \mathbf{R}^d\colon\ C(t) = \int_{R^d} cos(t \bullet s) dF(s) \tag{2.1.3}$$

where $F(\cdot)$ is some positive and bounded measure on $\mathbf{R}^d$. That is, $C(\cdot)$ is the Fourier-transform of $F(\cdot)$ (the spectral measure).

Assume that $C(0) > 0$. Then, with $\Phi \sim R([0, 2\pi[)$ and $\Omega \sim F/C(0)$, Y given by

$$Y(s) = \sqrt{2}\cos(\Phi + \Omega \bullet s),$$

is a second order stationary process with covariogram $C(\cdot)$ (see the section on the spectral method in Chapter 3). Here $\bullet$ denotes the usual inner product in $\mathbf{R}^n$

The definition V2 can be justified in a similar way because we have equivalence of the following two statements:

c) $2\gamma(\cdot)$ is the variogram of an intrinsically stationary stochastic process Z.
d) $2\gamma(0) = 0$, $2\gamma(\cdot)$ is symmetric and $-2\gamma(\cdot)$ is conditionally nonnegative definite.

That c) implies $2\gamma(0) = 0$ and symmetry of $2\gamma(\cdot)$ i follows because $2\gamma(h)$ is the variance of the increments $Z(s+h) - Z(s)$ of Z for $h \in \mathbf{R}^n$. From Section 1 of Chapter 1 we have that the process Z is an IRF-0, and for

$$Y(s) = \sum_{i=1}^{n} a_i Z(s_i + s)$$

where $\sum_{i=1}^{n} a_i = 0$, the variance of $Y(s)$, $sinD$ is

$$Var(Y(s)) = -\sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j \gamma(s_i - s_j),$$

so that $-2\gamma(\cdot)$ is conditionally nonnegative definite.

To show that d) implies c), let for some fixed but arbitrary $s_0 \in \mathbf{R}^d$, the function $C(\cdot,\cdot)$ be given by

$$C(s,t) = \gamma(s - s_0) + \gamma(t - s_0) - \gamma(s - t)$$

Then $C(\cdot,\cdot)$ is nonnegative definite: With $(a_1,\ldots,a_n) \in \mathbf{R}^n$, $s_1,\ldots,s_n \in \mathbf{R}^d$ and $a_0 = -\sum_{i=1}^n a_i$,

$$\sum_{i=1}^n \sum_{j=1}^n a_i a_j C(s_i,s_j) = \sum_{i=1}^n \sum_{j=1}^n a_i a_j (\gamma(s_i - s_0) + \gamma(s_j - s_0) - \gamma(s_i - s_j)) =$$

$$\sum_{i=0}^n \sum_{j=0}^n a_i a_j (\gamma(s_i - s_0) + \gamma(s_j - s_0) - \gamma(s_i - s_j)) = -\sum_{i=0}^n \sum_{j=0}^n a_i a_j (\gamma(s_i - s_j)) \geq 0$$

As before there exists some Gaussian process Z with covariance function given by $C(\cdot,\cdot)$. The variance of $Z(s+h) - Z(s)$ is easily seen to be $2\gamma(h)$ so that Z is intrinsically stationary with variogram $2\gamma(\cdot)$.

2.2 Criteria of validity

The definitions V1–V2 are usually not very operational, as may be difficult in a direct way from the definitions to decide whether some function is nonnegative definite or conditionally nonnegative definite. In the next paragraphs different criteria for nonnegative definiteness or conditionally nonnegative definiteness of a given function will be described.

We consider first some criteria of validity which can be applied for functions proposed as covariograms.

First criteria of validity (Bochners theorem) According to Bochners theorem we have for $C(\cdot)$ real and continuous, that $C(\cdot)$ is nonnegative definite if and only if $C(\cdot)$ has the representation (2.1.3). If the spectral measure $F(\cdot)$ has a density $f(\cdot)$ then $f(\cdot)$ is given by

$$f(t) = \frac{1}{(2\pi)^n} \int_{\mathbf{R}^n} \cos(s \bullet t) C(s) ds$$

We can thus try to calculate $f(\cdot)$. If $f(\cdot)$ is nonnegative and integrable it follows that $C(\cdot)$ is a valid covariogram.

The following criteria is sufficient but not necessary for $C(\cdot)$ to be a valid covariogram.

Second criteria of validity This criteria of validity is due to Christakos (1984). A wide class of isotropic covariance models on $\mathbf{R}^d$ can be expressed as

$$C(r) = c \int_r^\infty u^d H(r/u) dP(u) \tag{2.2.4}$$

where

$$H(r/u) = B\left(\frac{d+1}{2}, \frac{1}{2}\right) \int_{\frac{r}{u}}^{1} \left(1 - v^2\right)^{(d-1)/2} dv$$

$B(\cdot,\cdot)$ is the beta function and $P(\cdot)$ is a positive measure (see Matérn, 1960). The second criteria of validity is obtained by assuming that P is absolutely continuous with respect to the Lebesgue measure, solving (2.2.4) with respect to dP/du, and finding the solutions for which $dP/du \geq 0$. We then get the following results for $d \leq 3$:
$C(\cdot)$ is a valid covariogram if

$d = 1 : C''(\cdot) > 0$
$d = 2 :$

$$\forall r \geq 0 : \int_{r}^{\infty} u\left(u^2 - r^2\right)^{-\frac{1}{2}} dC''(u) \geq 0$$

$d = 3 :$

$$\forall r \geq 0 : \; C''(r) - rC'''(r) \geq 0$$

I will only show the result in the case $d = 1$:
We let $f = dP/du$. Then

$$\begin{aligned} C(r) &= c \int_{r}^{\infty} uf(u) \int_{r/u}^{1} dvdu = c \int_{r}^{\infty} uf(u)(1 - r/u)du \\ &= c\left(\int_{r}^{\infty} uf(u)du - r \int_{r}^{\infty} f(u)du \right) = c\left([F_1(u)]_r^{\infty} - r[F_2(u)]_r^{\infty}\right) \\ &\Rightarrow C''(r) = cf(r) \end{aligned}$$

as the derivatives of $F_1(r)$ and $F_2(r)$ are $rf(r)$ and $f(r)$, respectively. That is, $f(r) = C''(r)/c$ is a solution to (2.2.4). The constant c is positive, so $C(\cdot)$ is a valid covariogram when $C''(u) > 0 \; \forall u > 0$.

If we wish to apply the second criteria of validity a necessary condition is that

$$C'(0) < 0.$$

We get this from (2.2.4) as follows ($d = 1$):

$$C'(0) = \lim_{r \to 0} c\left(\int_{r}^{\infty} uH(r/u)dP(u) - \int_{0}^{\infty} uH(0)dP(u) \right) /r =$$

$$\lim_{r\to 0} c\left(\int_r^\infty \frac{u}{r}(H(r/u)-H(0))dP(u) - \int_0^r \frac{u}{r}H(0)dP(u)\right)$$

The integral

$$\int_0^r \frac{u}{r}H(0)dP(u)$$

is positive and since $H(r/u) - H(0) = -2\int_0^{r/u} dv = -2r/u$

$$c\lim_{r\to 0}\int_r^\infty \frac{u}{r}(H(r/u)-H(0))dP(u) = -2c\int_0^\infty dP(u) < 0$$

It thus follows that $C'(0) < 0$.

Third criteria of validity (Schönberg, 1938)) This criteria can be applied for functions $C : \mathbf{R}^d \to \mathbf{R}$ with $C(s) = C^0(\|s\|)\ \forall s \in \mathbf{R}^d$ (i.e. isotropic functions). It then holds (Schönberg, 1938) that $C(\cdot)$ is nonnegative definite $\forall n = 1, 2, \ldots$ if and only if $C^0(\|\cdot\|)$ is completely monotone. A function $f : \mathbf{R} \to \mathbf{R}$ is completely monotone if $f(0) = f(0+)$ and $(-1)^n f^n(t) \geq 0$ for $0 < t < \infty$ and $n = 0, 1, 2, \ldots$.

The first, second and third criterion of validity can also be applied to functions proposed as variogram models for second order stationary processes because in this case the variogram $2\gamma(\cdot)$ is related to the covariogram $C(\cdot)$ by $2\gamma(h) = C(0) - C(h)$. The fourth criterion can be applied to check the validity of a function as a variogram for an intrinsically but not second order stationary process.

Fourth criterion of validity The fourth criterion of validity is based on an analogue of Bochners theorem. According to Matheron (1973) we have for $\gamma : \mathbf{R}^d \to \mathbf{R}$ continuous and symmetric that $\gamma(\cdot)$ is nonnegative definite if and only if

$$\forall s \in \mathbf{R} :\ \gamma(s) = \int_{\mathbf{R}^d} \frac{1-\cos(w \bullet s)}{\|w\|^2} G(dw) + K_0(s) \tag{2.2.5}$$

where $G(\cdot)$ is a nonnegative measure which is continuous at the origin and such that $\int \left(1+\|w\|^2\right)^{-1} G(dw) < \infty$ and $K_0(\cdot)$ is a conditionally nonnegative definite polynomial of degree less or equal to four. Following Christakos (1984) we can also derive some sufficient criteria from (2.2.5) for $\gamma(\cdot)$ to be a valid variogram. These sufficient criteria are:

a) $\gamma(\cdot)$ has a spectral representation

$$\gamma(s) = \int_{\mathbf{R}^d} \cos(w \bullet s)\Gamma(w)dw$$

where $\Gamma(\cdot) \leq 0$ and continuous at the origin.
b) $\lim_{||s|| \to \infty} \frac{\gamma(s)}{||s||^2} = 0$

Proof: By a) and b) it follows that with e an unit vector in $\mathbf{R}^d$ and $h > 0$ sufficiently large:

$$\gamma(eh) = \int_{\mathbf{R}^d} \cos(w \bullet eh)\Gamma(w)dw \leq ah^2$$

for some $a \geq 0$. The function $f(h) = h^2$ induces a measure on $\mathbf{R}$ which for $t < 0$ has the Laplacetransform $2/t^2$. Let e_i be the vector in $\mathbf{R}^d$ for which the i'th coordinate is one, and all other coordinates are zero. Then with $w_i = w \bullet e_i$,

$$\int_0^\infty e^{th} d\left(\int_{\mathbf{R}^d} \cos(e \bullet wh)\Gamma(w)dw\right) = -\int_0^\infty e^{th}\left(\int_{\mathbf{R}^d} \sin(hw_i)w_i\Gamma(w)dw\right)dh =$$
$$-\int_{\mathbf{R}^d} w_i\Gamma(w)\left(\int_0^\infty \sin(hw_i)e^{th}dh\right)dw = -\int_{\mathbf{R}^d} \frac{w_i^2}{t^2 + w_i^2}\Gamma(w)dw$$

where the first and second equality follows by Tonelli-Fubinis theorem. Summing over all $i = 1, \ldots, n$ we get

$$\sum_{i=1}^d -\int_{\mathbf{R}^d} \frac{w_i^2}{t^2 + w_i^2}\Gamma(w)dw \leq ad2/t^2 \Rightarrow \int_{\mathbf{R}^d} \sum_{i=1}^d \frac{w_i^2 t^2}{t^2 + w_i^2}(-\Gamma(w))dw \leq 2ad \Rightarrow$$
$$\int_{\mathbf{R}^d} -||w||^2\Gamma(w)dw \leq 2ad < \infty$$

(By letting $t \to -\infty$ and monotone convergence theorem). We can now write

$$\gamma(h) = \int_{\mathbf{R}^d} (1 - \cos(w \bullet h))(-\Gamma(w))dw - \int_{\mathbf{R}^d} (-\Gamma(w))dw$$

Since $\gamma(0) = 0$ the last integral is zero and hence

$$\gamma(h) = \int_{\mathbf{R}^d} \frac{(1 - \cos(w \bullet h))}{||w||^2}\left(-||w||^2\right)\Gamma(w)dw$$

It follows that $\gamma(\cdot)$ is of the form (2.2.5) so that $\gamma(\cdot)$ is a valid variogram. □

The condition b) is in fact necessary (Matheron, 1973, Section 2.2).

2.3 Construction of isotropic covariogram and variogram models

From the first criteria of validity it is clear that any Fourier-transform of a positive bounded measure will yield a valid continuous covariogram. Similarly, any function $\gamma(\cdot)$ obtained from the expression (2.2.5) is a valid variogram. We can e.g. consider the isotropic covariogram $C(\cdot)$ obtained as the characteristic function of $N_d(0, I)$:

$$\forall h \in \mathbf{R}^d : \; C(h) = \int_{\mathbf{R}^d} \exp(ih \bullet w) \exp\left(-||w||^2/2\right) dw = \exp\left(-||h||^2/2\right)$$

Every covariogram obtained from an isotropic spectral distribution is isotropic. This is easily seen because if $G(\cdot)$ is isotropic then for every orthogonal matrix A

$$C(h) = \int_{\mathbf{R}^d} \cos(w \bullet h) G(dw) = \int_{\mathbf{R}^d} \cos(wA \bullet hA) G(dwA) =$$
$$\int_{\mathbf{R}^d} \cos(hA \bullet v) G(dv) = C(hA)$$

Another way to construct an isotropic covariogram on $\mathbf{R}^d$ is by considering a covariogram $C_1(\cdot)$ on $\mathbf{R}$ and then let

$$C(h) = E(C_1(h \bullet \Theta))$$

where Θ is a stochastic vector uniformly distributed on the unit sphere. The function $C(\cdot)$ is then the covariogram of some stochastic process on $\mathbf{R}^d$ and is given by the formula

$$C(r) = \frac{\Gamma\left(\frac{d}{2}\right)}{\sqrt{\pi}\Gamma\left(\frac{d-1}{2}\right)} \int_{-1}^{1} C_1(vr)\left(1 - v^2\right)^{(d-3)/2} dv$$

(see the section on the turning bands method in Chapter 3). The covariogram $C_1(\cdot)$ also has a spectral representation

$$C_1(r) = \int_0^\infty \cos(rs) F(ds) \tag{2.2.6}$$

By combining these formulae and using a series-expansion of cos and a look-up in a table of integrals, we obtain that $C(\cdot)$ can be written in the form

$$C(r) = 2^{d/2-1}\Gamma(d/2) \int_0^\infty (sr)^{1-d/2} J_{d/2-1}(rs) F(ds) \tag{2.2.7}$$

where $J_{d/2-1}(\cdot)$ is the Bessel function of the first kind of order $d/2 - 1$. This is the spectral representation of an isotropic covariogram because the Fourier-transform of an

isotropic covariogram is a Hankel-transform. We thus see that for every covariogram $C_1(\cdot)$ there corresponds a covariogram $C(\cdot)$ on $\mathbf{R}^d$, and on the other hand: for every isotropic and continuous covariogram $C(\cdot)$ on $\mathbf{R}^d$ there corresponds a uniquely determined covariogram $C_1(\cdot)$ given as the Fourier-transform of the spectral measure $F(\cdot)$ related to $C(\cdot)$ through (2.2.7).

3 Simulation of Gaussian random fields

Simulation of a stochastic process (random field) can be useful for many purposes. One application is model identification where comparisons between the actual data and simulated realizations of a proposed model may help to choose a suitable model. Another important application is Monte Carlo estimation of properties concerning a stochastic process when these properties can not be evaluated analytically.

In this chapter different methods for simulation of a second order Gaussian random field $Z = \{Z(s) : s \in D\}$, $D \subseteq \mathbf{R}^d$ will be considered. In practice D is bounded and will be approximated by a finite set of sites $s_1, \ldots, s_n \in D$, where these sites could e.g. constitute a rectangular grid. We then wish to simulate the random vector $(Z(s_1), \ldots, Z(s_n))$ which has a multidimensional Gaussian distribution $N_n(\mu, \Sigma)$ where $\Sigma = (C(s_i - s_j))_{i,j=1,\ldots,n}$ and $C(\cdot)$ is the covariogram of Z. If n is small, methods based on the Cholesky decomposition of Σ can be applied, but such methods are computationally prohibitive when n is large. In the following I will describe four methods which can be applied for large n. In all cases we assume that $\mu = 0$.

3.1 The spectral method

This method can be applied when $C(\cdot)$ is continuous and $C(0) > 0$ ($C(0)$ is of course always different from zero in cases of interest). Then $C(\cdot)$ has the spectral representation (2.1.3) and the normed spectral measure $F(\cdot)/C(0)$ is a probability distribution. With $\Omega \sim F(\cdot)/C(0)$ and $\Phi \sim R([0, 2\pi[)$ independent, the random field

$$Y(s) = \sqrt{2C(0)}\cos(\Omega \bullet s + \Phi),\ s \in D \tag{3.1.1}$$

has $C(\cdot)$ as its covariogram.
Proof:

$$\phi(t) = E(\cos(\Omega \bullet s + \Phi)|\Omega = t) = E(\cos(s \bullet t + \Phi)) = E(\cos(\Phi)) = 0$$

so

$$E(\cos(\Omega \bullet s + \Phi)) = E(E(\cos(\Omega \bullet s + \Phi)|\Omega)) = E(\phi(\Omega)) = 0$$

and hence,

$$E(Y(s)Y(s+h)) = 2C(0)E(\cos(\Omega \bullet s + \Phi)\cos(\Omega \bullet (s+h) + \Phi)) =$$
$$2C(0)E\left(\frac{1}{2}(\cos(-\Omega \bullet h) + \cos(2(\Omega \bullet s + \Phi) + \Omega \bullet h))\right) =$$
$$C(0)E(\cos(-\Omega \bullet h)) + C(0)E(\cos(2(\Omega \bullet s + \Phi) + \Omega \bullet h)) = C(0)E(\cos(-\Omega \bullet h)) = C(h)$$

where we have used (2.1.3) in the last identity. □

If we average a sample $Y_1, \ldots, Y_n$ of independent realizations of Y then by the central limit theorem the resulting stochastic process

$$Z' = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} Y_i$$

has a distribution which is approximately Gaussian, and Z' has the covariogram $C(\cdot)$.
Example: For the exponential covariogram $C(h) = C(0)\exp(-\|h\|/a)$, $a \geq 0$, the spectral density is given by

$$f(s) = \frac{C(0)a^3}{\pi^2\left(1 + a^2||s||^2\right)}, \; s \in \mathbf{R}^d$$

which is the density of a Cauchy-distribution (Lantuéjoul, 1994).

3.2 The turning bands method

The turning bands method (Matheron, 1973) can be applied for simulation of an isotropic Gaussian random field by simulating anothe stochastic process on a one dimensional domain. The random field is isotropid if $C(h) = C^0(\|h\|)$, $h \in \mathbf{R}^d$. Let $C_1 : \mathbf{R} \to \mathbf{R}$ be the covariance function determined by

$$C^0(r) = \frac{\Gamma\left(\frac{d}{2}\right)}{\sqrt{\pi}\Gamma\left(\frac{d-1}{2}\right)} \int_{-1}^{1} C_1(vr)\left(1 - v^2\right)^{(d-3)/2} dv. \tag{3.2.1}$$

As stated in Section 3 of Chapter 2, (3.2.1) is the general form of an isotropic continuous covariance function on $\mathbf{R}^d$ so the covariance function $C_1(\cdot)$ corresponding to $C^0(\cdot)$ always exists and is given by (2.2.6) where F is the measure related C^0 by (2.2.7). If Y_1 is a zero-mean second order stationary process on $\mathbf{R}$ with $C_1(\cdot)$ as covariance function and if $\Theta = (\Theta_1, \ldots, \Theta_d)$ is uniformly distributed on the unit sphere in $\mathbf{R}^d$ and independent of Y_1, then the stochastic process Y given by

$$Y(s) = Y_1(\Theta \bullet s)$$

is a zero-mean second order stationary process with $C(\cdot)$ as its covariance function.
Proof: Note first that

$$E(Y(s)) = E(E(Y_1(\Theta \bullet s) \mid \Theta)) = 0$$

Let further $e_1 = (1, 0, \ldots, 0) \in \mathbf{R}^d$. Then

$$E(Y(s)Y(s+h)) = E(E(Y_1(\Theta \bullet s)Y_1(\Theta \bullet (s+h)) \mid \Theta)) =$$

$$E(C_1(\Theta \bullet h)) = E\left(C_1\left(\Theta \bullet \frac{h}{||h||}||h||\right)\right) = E(C_1(||h||\Theta \bullet e_1)) = E(C_1(\Theta_1||h||)) =$$

$$\int_{-1}^{1} \frac{\Gamma\left(\frac{d}{2}\right)}{\sqrt{\pi}\Gamma\left(\frac{d-1}{2}\right)}\left(1 - v^2\right)^{(d-3)/2} C_1(v||h||)dv = C^0(||h||).$$

The fourth equality is due to the rotation invariance of the distribution of Θ, and the second last equality holds because the density of Θ_1 is

$$\frac{\Gamma\left(\frac{d}{2}\right)}{\sqrt{\pi\Gamma\left(\frac{d-1}{2}\right)}}\left(1-v^2\right)^{(d-3)/2}, \quad -1 \leq v \leq 1$$

□

As for the spectral method we have to average independent realizations of Y in order to obtain a simulation of a stochastic process which is (approximately) Gaussian and has the covariogram $C(\cdot)$.

Practical experience (Lantuéjoul, 1994) shows that averaging a sample of Y corresponding to a fixed set of regularly dispersed directions in $\mathbf{R}^d$ is more efficient than averaging over the same number of randomly chosen directions.

For $d = 3$, $C_1(\cdot)$ is easy to find and it is given by

$$C_1(r) = r\frac{d}{dr}C^0(r)$$

The simulation of Y_1 may e.g. be performed by the spectral method or the dilution method but other alternatives exist (see Lantuéjoul, 1994).

3.3 The dilution method

The dilution method can be applied if the covariogram is of the form

$$C(h) = \lambda \int_{\mathbf{R}^d} g(x)g(x+h)dx \tag{3.3.1}$$

for some real function g and a constant $\lambda > 0$. Let

$$Y(s) = \sum_{u\in\Phi} \epsilon(u)g(x-u) \tag{3.3.2}$$

where Φ is a homogenuous Poisson point process in $\mathbf{R}^d$ with intensity λ and $\epsilon = (\epsilon_s)_{s\in\mathbf{R}^d}$ is a second order stochastic process where the $\epsilon_s's$ are iid with zero mean and variance equal to one. Then Y is a zero-mean stochastic process with covariogram $C(\cdot)$.

Proof: The mean of Y is zero because ϵ is zero-mean, and

$$E(Y(s)Y(s+h)) = E\left(E\left(\sum_{u\in\Phi,\, v\in\Phi} \epsilon(u)\epsilon(v)g(s-u)g(s+h-u) \mid \Phi\right)\right) =$$

$$E\left(\sum_{u\in\Phi} g(s-u)g(s+h-u)\right) = \lambda\int_{\mathbf{R}^d} g(s-u)g(s+h-u)du = \lambda\int_{\mathbf{R}^d} g(v)g(v+h)dv$$

The second last equality follows by Slivnyaks theorem (see e.g. Stoyan et al., 1987, p. 50). □

To obtain a sample which is approximately Gaussian we again have to average independent realizations of Y.

A similar approach is described in Ripley (1987). Ripley (1987) considers a stochastic process X on $\mathbf{R}$, given as a stochastic integral

$$X(s) = \int g(x-y)dW(y) \tag{3.3.3}$$

where W is a standard Wiener-process. By Yaglom (1960, p.451–453) any second order stationary process to which there corresponds a spectral density $f(\cdot)$ can be represented in this way where $g(\cdot)$ is the Fouriertransform of any function $h(\cdot)$ satisfying $|h(\lambda)|^2 = f(\lambda)$. The stochastic integral (3.3.3) is well-defined as the limit in mean square

$$X(s) = \lim_{N\to\infty}\lim_{n\to\infty}\sum_{j=0}^{n-1} g\left(s - N + j\frac{2N}{n}\right)\sqrt{\frac{2N}{n}}\epsilon_j, \quad s \in \mathbf{R}, \tag{3.3.4}$$

where $(\epsilon_j)_{j=1,\ldots,n}$ is a sequence of independent standard Gaussian random variables (see Adler, 1981, p. 28). Using (3.3.4) and the Cauchy-Schwarz inequality it follows that the covariance function of X is given by

$$C(h) = \int_{\mathbf{R}} g(x+h)g(x)dx, \quad h \in \mathbf{R}.$$

An approximate simulation of X is obtained using (3.3.4). That is,

$$X(s) \approx \sum_{j=0}^{n-1} g\left(s - N + j\frac{2N}{n}\right)\sqrt{\frac{2N}{n}}\epsilon_j \tag{3.3.4}$$

for large N and n.

For the exponential covariogram $g(\cdot)$ is given by

$$g(x) = \frac{1}{2\sqrt{\pi a}}\exp\left(-\frac{|x|}{a}\right)\frac{|x|}{a}$$

while for the spherical covariogram

$$C(h) = \left(1 - \frac{3}{2}\frac{|h|}{a} + \frac{1}{2}\frac{|h|^3}{a^3}\right)1(|h| \le a), \quad a > 0,$$

$g(\cdot)$ is given by

$$g(s) = s1_{\{|x|\le a\}}$$

(see Lantuéjoul, 1994)

It is in practice preferable that g has a bounded support so that (3.3.2) can be approximated by a finite sum (and the approximation (3.3.4) works well). Therefore the dilution method is not recommendable for an exponential covariogram while it may be suitable for the spherical covariogram.

3.4 Comparison of the spectral method, the turning bands method and the dilution method

The use of the dilution method is restricted by the assumption that the covariance is of the form (3.3.1), where it is in practice required that g has a bounded support. In contrary, the spectral method is quite general requiring only that the covariance function is continuous. The turning bands method seems less general requiring also isotropy, but a version of the turning bands method for nonisotropic covariograms has been developed (Mantouglou & Wilson, 1982). We can rewrite $Y(s)$ from (3.1.1) as

$$Y(s) = X(\Theta \bullet s)$$

where

$$X(t) = \sqrt{2C(0)}\cos(Rt + \Phi)$$

and (R, Θ) is the polar coordinates of Ω. An advantage of the turning bands method compared to the spectral method is that Y_1 can be chosen to be ergodic, while X is in general not ergodic. In Mantouglou & Wilson (1982) it is demonstrated, that the turning bands method is as accurate but less computationally expensive than the spectral method.

For the three methods mentioned an important question is how big samples should be averaged in order to approximate well enough the Gaussianity of the process subject for simulation. Attemps to answer this question via the Berry-Esseén theorem has been described in Lantuéjoul (1994), but the sample sizes needed according to these results are much bigger than practical experience suggests. A minimum sample size is 15 (Lantuéjoul, 1994).

3.5 Sequential simulation

A sequential simulation of a random vector $Z = (Z_1, \ldots, Z_n) = (Z(s_1), \ldots Z(s_n)$ can be performed by initially simulating a value z_1 of Z_1, then simulate a value of Z_2 from the conditional distribution of Z_2 given $Z_1 = z_1$ and soforth ending by simulation of Z_n given $Z_1 = z_1, \ldots, Z_{n-1} = z_{n-1}$.

When Z is Gaussian the conditional distributions are also Gaussian and the conditional expectations and variances can be calculated. When n is large however, approximations are called for in order to reduce the cost of computing the parameters in the conditional distributions. This can be done by reducing the number of conditioning values, that is, approximate the conditional density $f(z_i \mid z_1, \ldots, z_{i-1})$ by $f(z_i \mid z_j,\ j \in A)$ where A is a subset of $\{z_1, \ldots z_{i-1}\}$. One way to choose A is (Gómez-Hernandéz & Cassigara, 1994) to define an upper limit N_c for the cardinality of A and then choose A to be the (up to) N_c members of $\{z_1, \ldots z_{i-1}\}$ corresponding to sites closest to s_i in some sense. Often the distance is measured by the variogram. Due to the approximation applied it its preferable to choose the order in which $Z_1, \ldots, Z_n$ are simulated in a random way.

The search for the conditioning values can be quite expensive on a large grid and Gómez-Hernández & Cassigara (1994) therefore proposes that this search is only performed within a search neighbourhood around the site where simulation is to be performed. This means that even if more than N_c values have been simulated we sometimes condition on less than N_c values hoping that values outside the search neighbourhood does not affect the conditional distribution of the actual random variable much. The size of the search neigbourhood can e.g. be determined by the range of the variogram.

Zonal anisotropy is said to occur when the range of the variogram is infinite in some direction. Then the search neighbourhood is as large as the grid $(s_1, \ldots, s_n)$ in that direction and computational problems persist. A way to deal with such a problem is a multigrid approach where simulation is first performed on a coarse grid defined by a subset of $(s_1, \ldots, s_n)$ such that a large search neighbourhood contains a reasonable number of sites to be inspected in the coarse grid. Subsequently values corresponding to the rest of the sites is to be simulated and hopefully we can now on the coarse grid find N_c conditioning values not very far away from the actual site. Of course a larger number of grids of decreasing coarseness may be used.

The screening sequential algorithm (SSA) In Omre, Sølna & Tjelmeland (1992) A is chosen in a different way and thereby the screening sequential simulation algorithm is obtained. The set A is chosen in a way inspired by a certain Markov property called the pseudo-Markov property (Adler, 1981). A short description of this Markov property will be given. Consider X a stochastic process on $\mathbf{R}^d$ and an infinitely often differentiable closed surface ∂D which divides $\mathbf{R}^d$ into a bounded part D^- and an unbounded part D^+. The surface measure on ∂D is denoted $d\sigma$ and $\dot{t}$ is the unit normal vector to ∂D at the point $t \in \partial D$. For every function f with $supp(f) \subseteq \partial D$ compact and $\int_{\partial D} |f(t)|^2 dt < \infty$ we define

$$F(h) = \int_{\partial D} f(t) X\big(t + h\dot{t}\big) d\sigma(t), \; h \in \mathbf{R}$$

and

$$\Sigma_1 = \sigma\big(F'(0) : f \text{ as above}\big)$$

A σ-algebra Σ is called a splitting field for ∂D if the conditional independence property

$$\mathcal{D}\big(X(t) : t \in D^-, X(t) : t \in D^+ \mid \Sigma\big) = \mathcal{D}\big(X(t) : t \in D^- \mid \Sigma\big)\mathcal{D}\big(X(t) : t \in D^+ \mid \Sigma\big)$$

holds. If $F'(\cdot)$ is continuous at zero then X is pseudo-Markovian of order 1 if the minimal splitting field of ∂D is Σ_1. Inspired by this Markov property Omre, Sølna & Tjelmeland (1992) propose the following algorithm:

1) Define a set of directions

2) Simulate the values of the border of the grid
3) Repeat until a simulated value for every site in the grid is obtained:

> Choose a site inside the grid (e.g via some halfing procedure).
> Simulate a value from the conditional distribution, where the conditioning is on the two closest values on each side of the actual site in each of the specified directions.

The conditioning is on the two closest values because these values determine a difference quotient on each side of the actual site (with respect to each of the specified directions).

Evaluation of the sequential algorithms The screening sequential algorithms can not be justified by theoretical considerations and it must therefore be evaluated in an empirical manner. This also holds for the algorithms resulting from the approximations suggested by Gómez-Hernández & Cassigara (1994). The algorithms can be evaluated by estimating the finite-dimensional distributions from simulations obtained by the algorithms and then compare these estimated distributions with the true distributions. A less ambitious approach is to estimate only the mean, variogram, and univariate distribution from the simulations. The variogram is well reproduced from simulations obtained by the SSA and the simulation-algorithms proposed by Gómez-Hernández & Cassigara (1994) when attemps to simulate processes with exponential or doubly exponential variograms are made. This is not the case for the spherical variogram where the estimates are seriously biased. In Omre, Sølna & Tjelmeland (1992) no attemps are made to check the Gaussianity of the simulations while Gómez-Hernández & Cassigara (1994) asses the Gaussianity by considering histograms of the univariate distributions.

It is expected that the approximations to the sequential simulation method work well for the exponential and double-exponential variograms since second order stationary processes with these variograms posses certain Markov properties (Gómez-Hernandéz & Cassigara, 1994).

If the simulations are Gaussian (which is not well justified) the sequential simulation approaches proposed by Gómez-Hernández & Cassigara (1994) and Omre, Sølna & Tjelmeland (1992) has the advantage, compared to the three other simulation methods described in this chapter, that no averaging over independent samples is needed in order to obtain Gaussianity. On the other hand the approximations to the sequential simulation method which are proposed may only be suitable for simulation of processes with exponential or double exponential variograms.

Appendix Bochners theorem and the spectral representation of stationary processes

In this appendix a proof of Bochners theorem (Bochner, 1933) will be given in the onedimensional case. We state the theorem once more:

Theorem A continuous function $C : \mathbf{R}^n \to \mathbf{C}$ is non-negative definite if and only if it has the representation

A1 $$\forall t \in \mathbf{R}^d : \; C(t) = \int_{\mathbf{R}^n} \exp(it \bullet \lambda) dF(\lambda)$$

where $F(\cdot)$ is a bounded and positive measure.

Proof: ($n = 1$) The proof is due to Cramér (see Cramér & Leadbetter, 1967).

We consider first the "if"-part. Let $(a_1, \ldots, a_m)$ be an arbitrary vector in $\mathbf{C}^m$ not equal to zero. Then

$$\sum_{i=1}^{m}\sum_{j=1}^{m} a_i \bar{a}_j C(t_i - t_j) = \left\| \sum_{i=1}^{m} a_i g_i \right\|_2 = \geq 0$$

where $g_i(\lambda) = \exp(it_i\lambda)$ and $\| \cdot \|_2$ denotes $L_2(F)$ norm.

If on the other hand $C(\cdot)$ is nonnegative definite and continuous it follows for every function $f(\cdot)$ continuous on $]a, b[$ that

$$\int_a^b \int_a^b C(t-u) f(u) \overline{f(t)} dt du \geq 0$$

Then in particular for every $T > 0$ and $\lambda \in \mathbf{R}$,

$$g(\lambda, T) = \frac{1}{2\pi T} \int_0^T \int_0^T C(t-u) e^{-i\lambda(t-u)} dt du \geq 0$$

By a change of variable $z = t - u$ we obtain

$$g(\lambda, T) = \frac{1}{2\pi} \int_{-T}^{T} \left(1 - \frac{|z|}{T}\right) C(z) e^{-iz\lambda} dz = \frac{1}{2\pi} \int_{-\infty}^{\infty} h\left(\frac{z}{T}\right) C(z) e^{-iz\lambda} dz$$

where

$$h(x) = \begin{cases} 1 - |x| & |x| \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

We now want to show that $g(\lambda, T)$ is integrable over $\mathbf{R}$. Let $M > 0$, then we get

$$\int_{-\infty}^{\infty} h\left(\frac{\lambda}{2M}\right) g(\lambda, T) d\lambda = \frac{1}{2\pi} \int_{-\infty}^{\infty} \left(h\left(\frac{t}{T}\right) C(t) \int_{-\infty}^{\infty} h\left(\frac{\lambda}{2M}\right) e^{-it\lambda} d\lambda \right) dt =$$

$$\frac{1}{\pi} M \int_{-\infty}^{\infty} h\left(\frac{t}{T}\right) C(t) \left(\frac{\sin Mt}{t}\right)^2 dt \leq \frac{1}{\pi} C(0) \int_{-\infty}^{\infty} \left(\frac{\sin t}{t}\right)^2 dt = C(0)$$

By letting $M \to \infty$ we get by the monotone convergence theorem that

$$\int_{-\infty}^{\infty} g(\lambda, T) d\lambda \leq C(0)$$

The functions $g(\cdot, T)$ and $h(t/T)C(t)$ are both integrable and $g(\cdot, T)$ is the Fourier-transform of $h(t/T)C(t)$. By the inversion theorem for Fourier transforms

$$h\left(\frac{t}{T}\right) C(t) = \int_{-\infty}^{\infty} g(\lambda, T) e^{it\lambda} d\lambda$$

and for $t = 0$

$$C(0) = \int_{-\infty}^{\infty} g(\lambda, T) d\lambda$$

such that $h(|t|/T)C(t)/C(0)$ is the characteristic function of the density $g(\cdot, T)/C(0)$. If we let $T \to \infty$ then $h(t/T)C(t)/C(0) \to C(t)/C(0)$ which is continuous as a function of t. It follows that $C(t)/C(0)$ is the characteristic function of some distribution function $F(\cdot)$ and the theorem is proved. □

In analogy with the spectral representation of the covariance function given by Bochners theorem the second-order stationary processes themselves have a spectral representation which is fundamental to the study of such processes. This spectral representation is given by the following theorem (see e.g. Adler, 1981).

Theorem For every zero-mean, second order stationary process $(X(t))_{t \in \mathbf{R}^n}$ which is mean square continuous there exists a stochastic process $(Z(t))_{t \in \mathbf{R}^n}$ with orthogonal increments such that

A2 $$\forall t \in \mathbf{R}^n : \; X(t) \stackrel{a.s}{=} \int_{\mathbf{R}^n} \exp(it \bullet \lambda) dZ(\lambda)$$

Setting $Z(-\infty, \ldots, -\infty) = 0$ then Z is uniquely determined and

A3 $$E(Z(\lambda)) = 0, \;\; E\left(|Z(\lambda)|^2\right) = F(\lambda) \;, \;\; E\left(|Z(I)|^2\right) = F(I)$$

where I is any interval in $\mathbf{R}^n$ and $F(\cdot)$ is the spectral distribution function determined by A1 where $F((-\infty,\ldots,-\infty)) = 0$.

For an interval $I =]a_1, b_1] \times \cdots \times]a_n, b_n]$, $Z(I)$ is given by $Z(I) = Z(b_1,\ldots,b_n) - (Z(a_1,b_2,\ldots,b_n) + \cdots + Z(b_1,\ldots,b_{n-1},a_n)) + \cdots + (-1)^n Z(a_1,\ldots,a_n)$. The stochastic process Z then has orthogonal increments if for all disjoint intervals I_1 and I_2 in $\mathbf{R}^n$,

$$E\Big(Z(I_1)\overline{Z(I_2)}\Big) = 0$$

The integral on the right hand side of A2 is a mean-square integral (see e.g. Adler, 1981, p. 28) which is welldefined because

$$\int_{\mathbf{R}^n} |\exp(it \bullet \lambda)|^2 dF(\lambda) = \int_{\mathbf{R}^n} dF(\lambda) < \infty$$

Here only a brief survey of the proof of the theorem will be given. Only the case $n = 1$ is considered. The full proof can be found in Cramér & Leadbetter (1967). Let $H(X)$ denote the Hilbert space spanned by $X(t), t \in \mathbf{R}$. We now establish a correspondence between $H(X)$ and $L_2(F)$. First for every $t \in \mathbf{R}$ we let $X(t)$ and $\exp(it\lambda)$ be corresponding elements. By Bochners theorem inner products are preserved by this correspondence as

$$E\Big(X(t)\overline{X(u)}\Big) = \int_{-\infty}^{\infty} e^{it\lambda}\overline{e^{it\lambda}} dF(\lambda)$$

This implies also that the correspondence is one-to-one (with respect to the distance measures induced by the inner products on $H(X)$ and $L_2(F)$). The correspondence is now extended to all finite linear combinations of elements in $H(X)$ and $L_2(F)$ and limits of sequences of such linear combinations and a one-to-one correspondence between the whole of the two spaces is thus established. For every $\lambda_0 \in \mathbf{R}$ there to the indicator function $1(\cdot \leq \lambda_0)$ corresponds an element $Z(\lambda_0) \in H(X)$ and from the preservation of inner products it follows that $(Z(\lambda))_{\lambda\in\mathbf{R}}$ has orthogonal increments and that the relations A3 hold. Finally, let $\{T_n\}_{n\geq 1}$ be an increasing sequence of positive numbers such that $T_n \to \infty$ as $n \to \infty$ and let for each n, $-T_n = \lambda_1 < \lambda_2 < \cdots < \lambda_{m_n+1} = T_n$ be a subdivision of the interval $]-T_n, T_n[$ such that $\sup_{j=1,\ldots,m_n} |\lambda_{j+1} - \lambda_j| \to \infty$ as $n \to \infty$. The stochastic variable

$$Y_n = \sum_{j=1}^{m_n} e^{it\lambda_j}(Z(\lambda_{j+1}) - Z(\lambda_j))$$

corresponds to the stepfunction

$$g_n(\lambda) = \begin{cases} e^{it\lambda_j} & \lambda_j < \lambda \leq \lambda_{j+1}\ j = 1,\ldots,m_n \\ 0 & \text{otherwise} \end{cases}$$

and

$$g_n(\lambda) \to e^{it\lambda} \in L_2(F) \text{ and } Y_n \to \int_{-\infty}^{\infty} e^{it\lambda} dZ(\lambda) \in H(X)$$

as $n \to \infty$ (the last convergence in mean-square). Since $e^{it\lambda}$ corresponds to $X(t)$ and the correspondence is one-to-one the theorem is proved $\square$

References

Adler, R. (1981). *The Geometry of Random Fields*. Wiley.

Armstrong, M. (1984). "Problems with Universal Kriging". *Math. Geol.*, 16.

Bochner, S. (1933). "Monotone Funktionen Stieltjessche Integrale und Harmonische Analyse". *Math. Ann.*, 108.

Christakos, G. (1984). "On the problem of permissible covariance and variogram models". *Water Resources Research*, 20.

Christensen, R. (1990). "The Equivalence of Predictions from Universal Kriging and Intrinsic Random Function-Kriging". *Math. Geol.*, 22.

Cramér, H. & Leadbetter, M. (1967). *Stationary and related stochastic processes*. Wiley.

Cressie, N. (1991). *Statistics for Spatial Data*. Wiley.

Delfiner, P. (1976). "Linear prediction of nonstationary spatial phenomena". In Guarascio, M.; David, M. & Huijbregts, C., editors, *Advanced Geostatistics in the Mining Industry*. Dordrecht.

Gómez-Hernández, J. & Cassigara, E. (1994). "Theory and Practice of Sequential Simulation". In Armstrong, M. & Dowd, P., editors, *Geostatistical Simulations*. Kluwer Academic Publishers.

Journel, A. & Huijbregts, C. (1978). *Mining Geostatistics*. Academic Press, London.

Lantuéjoul, C. (1994). "Nonconditional simulation of stationary isotropic multigaussian random functions". In Armstrong, M. & Dowd, P., editors, *Geostatistical Simulations*. Kluwer Academic Publishers.

Mantouglou, A. & Wilson, J. (1982). "The turning bands method for simulation of random fields using line generation by a spectral method". *Water Resources Research*, 18.

Matheron, G. (1973). "The Intrinsic Random Functions and their Applications". *Adv. Appl. Prob.*, 5.

Omre, H.; Sølna, K. & Tjelmeland, H. (1992). "Simulation of random functions on large lattices". In *Proceedings from Fourth International Geostatistical Congress; Troia, Portugal.*

Ripley, B. (1987). *Stochastic simulation*. Wiley, New York.

Schönberg, I. (1938). "Metric spaces and completely monotone functions". *Annals of Mathematics*, 39.

Stein, M. (1987). "Minimum Norm Quadratic Estimation of Spatial Variograms". *J. Amer. Statist. Assoc.*, 82:765–772.

Stoyan, D.; Kendall, W. S. & Mecke, J. (1987). *Stochastic Geometry and its Applications*. Wiley/Akademie-Verlag, Berlin.

Yaglom, A. (1960). *Correlation Theory of Stationary and Related Random Functions*. Springer-Verlag.

$s_1 = 0 \quad s_2 = h$

$$Y(s) = Z(s_1 + s) - Z(s_2 + s) = Z(s) - Z(s+h)$$

$$Cov(Y(s), Y(t)) = E\left((Z(s) - Z(s+h))(Z(t) - Z(t+h)) \right)$$

$$= E\left(\left[Z(s) - Z(s_0) + Z(s_0) - Z(s+h) \right] \left[Z(t) - Z(s_0) + Z(s_0) - Z(t+h) \right] \right)$$

$$2\gamma(s-t) = Var\left[Z(s) - Z(t) \right] = E\left(\left[Z(s) - Z(s_0) + Z(s_0) - Z(t) \right]^2 \right)$$

$$= E\left[\left[Z(s) - Z(s_0) \right]^2 + \left[Z(s_0) - Z(t) \right]^2 - 2 \left(Z(s) - Z(s_0) \right)\left(Z(s_0) - Z(t) \right) \right]$$

$$= 2\gamma(s - s_0) + 2\gamma(t - s_0) - 2\, Cov\left(Z(s) - Z(s_0), Z(s_0) - Z(t) \right)$$

$$\Leftrightarrow Cov\left(Z(s) - Z(s_0), Z(t) - Z(s_0) \right) = -\gamma(s-t) + \gamma(s - s_0) + \gamma(t - s_0)$$

$$= Cov(s - s_0, t - s_0) + Cov(s - s_0, t + h - s_0) - Cov(s + h - s_0, t - s_0) +$$

$$Cov(s + h - s_0, t + h - s_0)$$

$$= -\gamma(s-t) + \gamma(s - s_0) + \gamma(t - s_0) + \gamma(s - t - h) - \gamma(s - s_0) - \gamma(t + h - s_0)$$

$$+ \gamma(s + h - t) - \gamma(s + h - s_0) - \gamma(t - s_0) + \gamma(s - t)$$

$$+ \gamma(s + h - s_0) + \gamma(t + h - s_0)$$

$$= -2\gamma(s-t) + \gamma(s - t - h) + \gamma(s - t + h)$$

i.e. depends on (s,t) only on $(s-t)$

$$\mathbb{E}\left(\left(\frac{Z(x+\frac{1}{n})-Z(x)}{\frac{1}{n}}-\frac{Z(x+\frac{1}{m})-Z(x)}{\frac{1}{m}}\right)^2\right)$$

$$=\mathbb{E}\left((m-n)Z(x)+nZ(x+\tfrac{1}{n})-Z(x+\tfrac{1}{m})m\right)^2$$

$$=\mathbb{E}\Big[(m-n)^2Z(x)^2+2n(m-n)Z(x)Z(x+\tfrac{1}{n})-2m(n-m)Z(x)Z(x+\tfrac{1}{m})$$
$$+n^2Z(x+\tfrac{1}{n})^2-2mnZ(x+\tfrac{1}{n})Z(x+\tfrac{1}{m})+m^2Z(x+\tfrac{1}{m})^2\Big]$$

$$=(m-n)^2C(0)+2n(m-n)C(\tfrac{1}{n})-2m(n-m)C(\tfrac{1}{m})$$
$$+n^2C(0)-2nnC(\tfrac{1}{n}-\tfrac{1}{m})+m^2C(0)$$

$$=2m^2C(0)+2n^2C(0)-2nmC(0)+2nmC(\tfrac{1}{n})+2n^2C(\tfrac{1}{n})$$
$$-2nmC(\tfrac{1}{m})+2m^2C(\tfrac{1}{m}) \qquad \to 4C'(0)$$

$$\mathbb{E}\quad n^2\left[C(0)-C(\tfrac{1}{n})\right]+m^2\left[C(0)-C(\tfrac{1}{m})\right]$$

$$-2\,\mathbb{E}\left[\left(Z(x+\tfrac{1}{n})-Z(x)\right)\left(Z(x+\tfrac{1}{m})-Z(x)\right)\right]mn$$

$$-2nnC(\tfrac{1}{n}-\tfrac{1}{m})+2mnC(\tfrac{1}{n})+2mnC(\tfrac{1}{m})-2C(0)mn$$

$n\to\infty \quad 2mC'(\tfrac{1}{n})-2n^2C'(0)$ $\qquad (\tfrac{1}{m})$

$$\to 2C''(0)$$