Spatio-temporal statistics (MATH4341) Michaelmas term, 2023

Handout 3: Point referenced data modeling / Geostatistics

Lecturer & author: Georgios P. Karagiannis georgios.karagiannis@durham.ac.uk

Aim. To introduce Point referenced data modeling / Geostatistics: regional variables, random field, variogram,

Reading list & references:

[1] Cressie, N. (2015; Part I). Statistics for spatial data. John Wiley & Sons.

[2] Gaetan, C., & Guyon, X. (2010; Ch 2 & 5.1). Spatial statistics and modeling (Vol. 90). New York: Springer.

Specialized reading.

[3] Wackernagel, H. (2003). Multivariate geostatistics: an introduction with applications. Springer Science & Business Media. (on Geostatistics)

[4] Kent, J. T., & Mardia, K. V. (2022). Spatial analysis (Vol. 72). John Wiley & Sons. (on Spatial analysis)

Part 1. Intro to building stochastic models & concepts

Note 1. We discuss basic stochastic models and concepts for modeling point referenced data in the Geostatistics framework.

1. Stochastic processes (or Random fields)

Definition 2. A stochastic process (or random field) $Z = (Z_s; s \in \mathcal{S})$ taking values in $\mathcal{Z} \subseteq \mathbb{R}^q$, $q \geq 1$ is a family of random variables $\{Z_s := Z_s(\omega); s \in \mathcal{S}, \omega \in \Omega\}$ defined on the same probability space $(\Omega, \mathfrak{F}, \Pr)$ and taking values in $\mathcal{Z}$. The label $s \in \mathcal{S}$ is called site, the set $\mathcal{S} \subseteq \mathbb{R}^d$ is called the (spatial) set of sites at which the process is defined, and $\mathcal{Z}$ is called the state space of the process.

Note 3. Given a set $\{s_1, ..., s_n\}$ of sites, with $s_i \in S$, the random vector $(Z(s_1), ..., Z(s_n))^\top$ has a well-defined probability distribution that is completely determined by its joint CDF

$$\mathrm{F}_{s_1,...,s_n}(z_1, ..., z_n) = \Pr(Z(s_1) \leq z_1, ..., Z(s_n) \leq z_n)$$

Finite dimensional distributions (or fidi's) of Z is called the ensemble of all such joint CDF's with $n \in \mathbb{N}$ and $\{s_i \in S\}$.

Note 4. According to Kolmogorov Thm 5, to define a random field model, one must specify the joint distribution of $(Z(s_1), ..., Z(s_n))^\top$ for all of n and all $\{s_i \in S\}_{i=1}^n$ in a consistent way.

Proposition 5. *(Kolmogorov consistency theorem) Let* $\Pr_{s_1,...,s_n}$ *be a probability on* $\mathbb{R}^n$ *with join CDF* $F_{s_1,...,s_n}$ *for every finite collection of points* $s_1, ..., s_n$*. If* $F_{s_1,...,s_n}$ *is symmetric w.r.t. any permutation* $\mathfrak{p}$

$$F_{\mathfrak{p}(s_1),...,\mathfrak{p}(s_n)}\left(z_{\mathfrak{p}(1)}, ..., z_{\mathfrak{p}(n)}\right) = F_{s_1,...,s_n}\left(z_1, ..., z_n\right)$$

for all $n \in \mathbb{N}$, $\{s_i \in S\}$, *and* $\{z \in \mathbb{R}\}$, *and all if all permulations* $\mathfrak{p}$ *are consistent in the sense*

$$\lim_{z_n \to \infty} F_{s_1,...,s_n}\left(z_1, ..., z_n\right) = F_{s_1,...,s_{n-1}}\left(z_1, ..., z_{n-1}\right)$$

or all $n \in \mathbb{N}$, $\{s_i \in S\}$, *and* $\{z \in \mathbb{R}\}$, *then there exists a random field* Z *whose fidi's coincide with those in* F*.*

Example 6. Let $n \in \mathbb{N}$, let $\{f_i : T \to \mathbb{R};\ i = 1, ..., n\}$ be a set of constant functions, and let $\{Z_i \sim \mathrm{N}(0, 1)\}_{i=1}^{n}$ be a set of independent random variables. Then

$$\tilde{Z}_s = \sum_{i=1}^{n} Z_i f_i(s), \quad s \in S \tag{1.1}$$

is a well defined stochastic process as it satisfies Thm 5.

1.1. Mean and covariance functions.

Definition 7. The mean function $\mu(\cdot)$ and covariance function $c(\cdot, \cdot)$ of a random field $Z = (Z_s)_{s \in S}$ are defined as

$$\mu(s) = \mathrm{E}(Z_s), \qquad \forall s \in S \tag{1.2}$$

$$c(s, s') = \mathrm{Cov}(Z_s, Z_{s'}) = \mathrm{E}\left((Z_s - \mu(s))(Z_{s'} - \mu(s'))^{\top}\right), \qquad \forall s, s' \in S \tag{1.3}$$

Example 8. For (1.1), the mean function is $\mu(s) = \mathrm{E}\left(\tilde{Z}_s\right) = 0$ and covariance function is

$$\begin{aligned} c(s, s') = \mathrm{Cov}(Z_s, Z_{s'}) &= \mathrm{Cov}\left(\sum_{i=1}^{n} Z_i f_i(s), \sum_{j=1}^{n} Z_j f_j(s')\right) \\ &= \sum_{i=1}^{n} f_i(s) \sum_{j=1}^{n} f_j(s') \overset{1(i=j)}{\cancel{\mathrm{Cov}(Z_i, Z_j)}} = \sum_{i=1}^{n} f_i(s) f_i(s') \end{aligned}$$

1.1.1. *Construction of covariance functions.*

(The following provides the means for checking and constructing covariance functions.)

Proposition 9. *The function* $c : S \times S \to \mathbb{R}$, $S \subseteq \mathbb{R}^d$ is the covariance function iff $c(\cdot, \cdot)$ is semi-positive definite; i.e. the Gram matrix $(c(s_i, s_j))_{i,j=1}^{n}$ is non-negative definite for any $\{s_i\}_{i=1}^{n}$, $n \in \mathbb{N}$.

Example 10. $c(s,s') = 1(s = s')$ is a proper covariance function as

$$\sum_i \sum_j a_i a_j c(s_i, s_j) = \sum_i a_i^2 \geq 0, \ \ \forall a$$

Note 11. Prop 12 uses the experience from basis functions, while Theorem 30 uses experience from characteristic functions to be incorporated int he process for modeling reasons.

Remark 12. One way to construct a c.f c is to set $c(s,s') = \psi(s)^\top \psi(s')$, for a given vector of basis functions $\psi(\cdot) = (\psi_1(\cdot), ..., \psi_n(\cdot))$.

Proof. From Prop 9, as

$$\sum_i \sum_j a_i a_j c(s_i, s_j) = (\psi a)^\top (\psi a) \geq 0, \ \ \forall a \in \mathbb{R}^n$$

$\square$

2. Second order processes (or Random fields)

Definition 13. Second order process (or random field) $Z = (Z_s; s \in \mathcal{S})$ is called the stochastic process where $\mathrm{E}(Z_s^2) < \infty$ for all $s \in S$. Then the associated mean function $\mu(\cdot)$ and covariance function $c(\cdot,\cdot)$ exist.

3. Gaussian process

Definition 14. $Z = (Z_s; s \in S)$ indexed by $S \subseteq \mathbb{R}^d$ is a Gaussian process (GP) or random field (GRF) if for any $n \in \mathbb{N}$ and for any finite set $\{s_1, ..., s_n; s_i \in \mathcal{S}\}$, the random vector $(Z_{s_1}, ..., Z_{s_n})^\top$ has a multivariate normal distribution.

Also Example of Proposition

Proposition 15. *A GP $Z = (Z_s; s \in S)$ is fully characterized by its mean function $\mu : S \to \mathbb{R}$ with $\mu(s) = E(Z_s)$, and its covariance function with $c(s,s') = Cov(Z_s, Z_{s'})$.*

Notation 16. Hence, we denote the GP as $Z(\cdot) \sim \mathcal{GP}(\mu(\cdot), c(\cdot,\cdot))$.

Example 17. When using the GP as a model we may need to parameterize its parameters. An example of mean functions are polynomial expansions, such as $\mu(s) = \sum_{j=0}^{p-1} \beta_j s^j$ for some tunable unknown parameter β. Some examples of covariance functions (c.f.), for some tunable unknown parameter β,σ^2 are

1. Exponential c.f. $c(s,s') = \sigma^2 \exp(-\beta \|s - s'\|_1)$
2. Gaussian c.f. $c(s,s') = \sigma^2 \exp\left(-\beta \|s - s'\|_2^2\right)$
3. Nugget c.f. $c(s,s') = \sigma^2 1(s = s')$

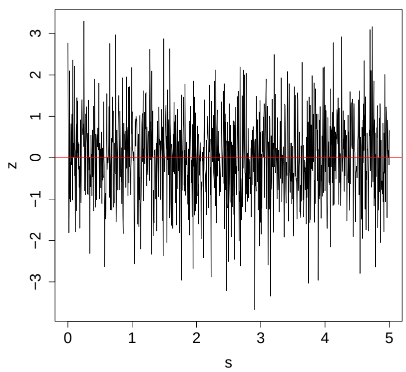
Example 18. Recall your linear regression lessons where you specified a sampling distribution $y_x|\beta, \sigma^2 \overset{\text{ind}}{\sim} \mathrm{N}\left(x^\top \beta, \sigma^2\right)$, $\forall x \in \mathbb{R}^d$; well that can be considered as a GP with $\mu_x = x^\top \beta$ and $c(x,x') = \sigma^2 1(x = x')$ in (3).

Example 19. Figs. 3.1 & 3.2 presents realizations of GRF $Z(\cdot) \sim \mathcal{GP}(\mu(\cdot), c(\cdot,\cdot))$ with $\mu(s) = 0$ and differently parameterized covariance functions in 1D and 2D. In 1D the code to simulate the GP is given in Algorithm 1. Note that we actually discretize it and simulate it from the fidi.

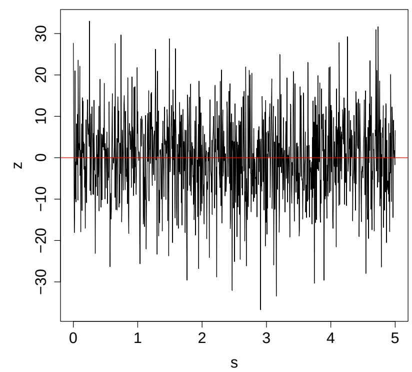
Algorithm 1 R script for simulating from a GP $(Z_s; s \in \mathbb{R}^1)$ with $\mu(s) = 0$ and $c(s,t) = \sigma^2 \exp\left(-\beta \|s-t\|_2^2\right)$

```
# set the GP parameterized mean and covariance function
mu_fun <- function(s) { return (0) }
cov_fun_gauss <- function(s,t,sig2,beta) { return (
sig2*exp(-beta*norm(c(s-t),type="2")**2) ) }
# discretize the problem in n = 100 spatial points
n <- 100
s_vec <- seq(from = 0, to = 5, length = n)
mu_vec <- matrix(nrow = n, ncol = 1)
Cov_mat <- matrix(nrow = n, ncol = n)
# compute the associated mean vector and covariance matrix of the n=100 dimensional
Normal r.v.
sig2_val <- 1.0 ;
beta_val <- 5
for (i in 1:n) {
    mu_vec[i] <- mu_fun(s_vec[i])
    for (j in 1:n) {
        Cov_mat[i,j] <- cov_fun_gauss(s_vec[i],s_vec[j],sig2_val,beta_val)
    }
}
# simulate from the associated distribution
z_vec <- mu_vec + t(chol(Cov_mat))%*%rnorm(n, mean=0, sd=1)
# plot the path (R produces a line plot)
plot(s_vec, z_vec, type="l")
abline(h=0,col="red")
```

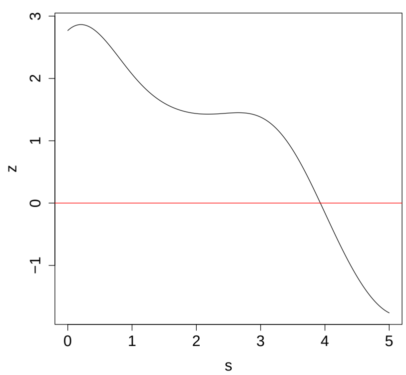
Nugget c.f. is the usual noise where the height of ups and downs are random and controlled by σ^2 (Fig. 3.1a & 3.1b ; Fig. 3.2a & 3.2b). In Gaussian c.f. the height of ups and downs are random and controlled by σ^2 (Fig.3.1c & 3.1d ; Fig. 3.2c & 3.2d), and the spatial dependence / frequency of the ups and downs is controlled by β (Fig. 3.1d & 3.1e ; Fig. 3.2d & 3.2e). Realizations with different c.f. have different behavior (Fig. 3.1a, 3.1d & 3.1e ; Fig. 3.2a, 3.2d & 3.2e)



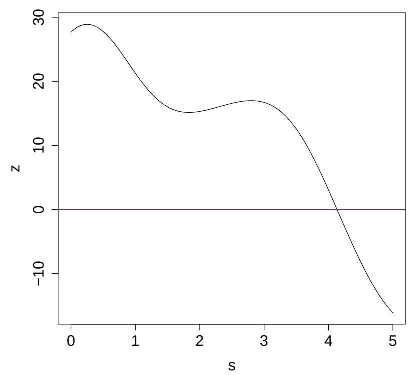
(A) Nugget c.f
($\sigma^2 = 1$)



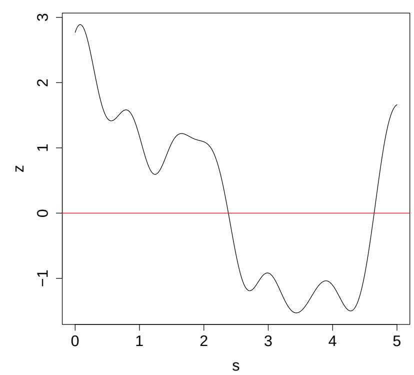
(B) Nugget c.f
($\sigma^2 = 100$)



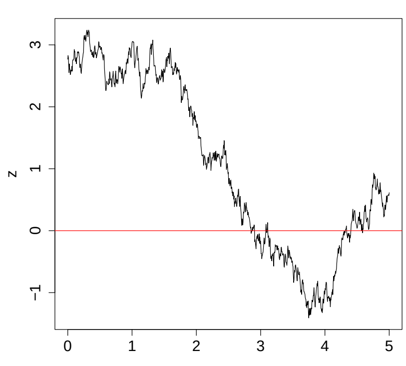
(C) Gauss c.f
($\sigma^2 = 1$,$\beta = 0.5$)



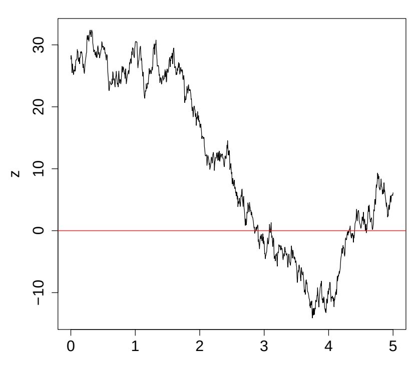
(D) Gauss c.f
($\sigma^2 = 100$,$\beta = 0.5$)



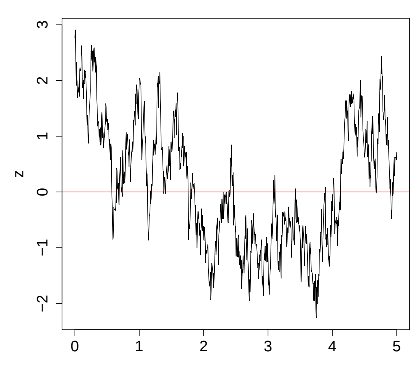
(E) Gauss c.f
($\sigma^2 = 1$,$\beta = 5$)



(F) Exp c.f
($\sigma^2 = 1$,$\beta = 0.5$)

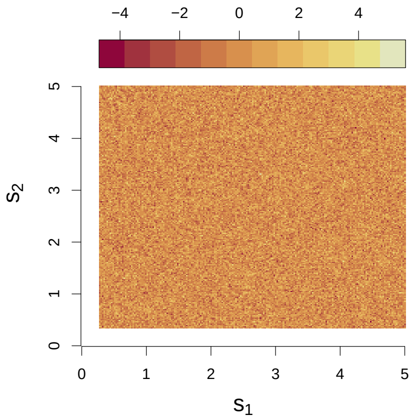


(G) Exp c.f
($\sigma^2 = 100$,$\beta = 0.5$)

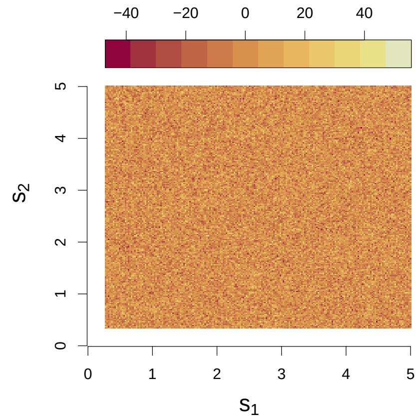


(H) Exp c.f
($\sigma^2 = 1$,$\beta = 5$)

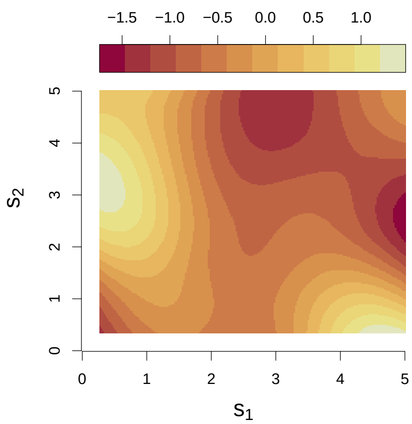
FIGURE 3.1. Realizations of GRF $Z(\cdot) \sim \mathcal{GP}(\mu(\cdot), c(\cdot,\cdot))$ when $s \in [0,5]$ (using same seed)



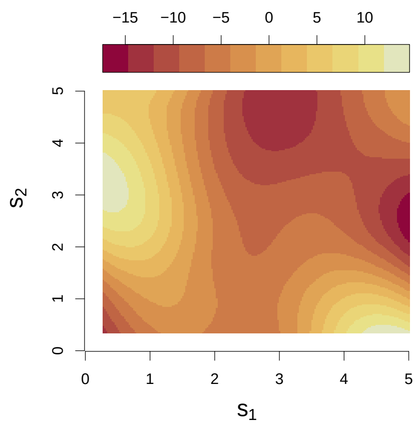
(A) Nugget c.f
($\sigma^2 = 1$)



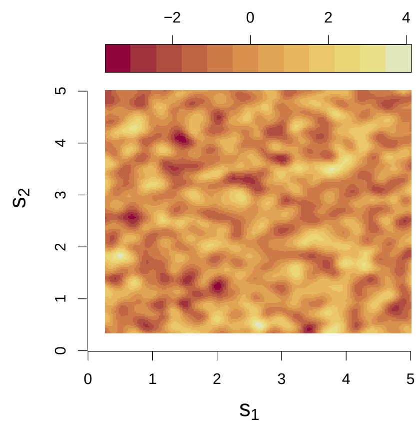
(B) Nugget c.f
($\sigma^2 = 100$)



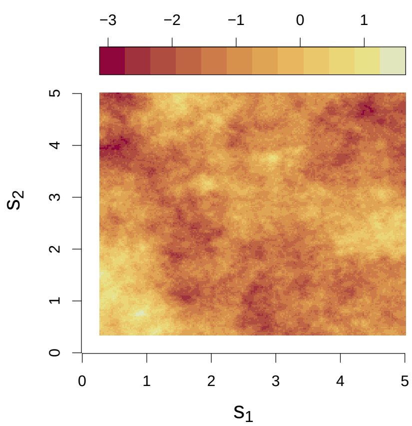
(C) Gauss c.f
($\sigma^2 = 1$,$\beta = 0.5$)



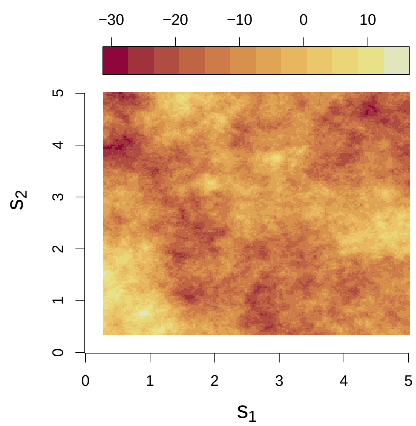
(D) Gauss c.f
($\sigma^2 = 100$,$\beta = 0.5$)



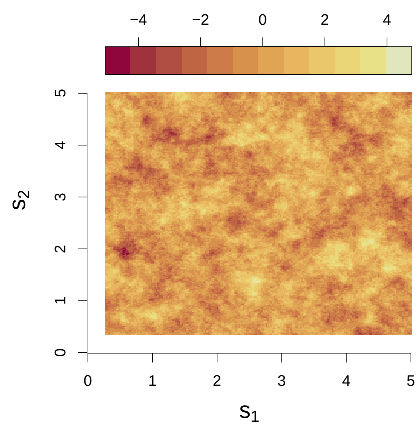
(E) Gauss c.f
($\sigma^2 = 1$,$\beta = 5$)



(F) Exp c.f
($\sigma^2 = 1$,$\beta = 0.5$)



(G) Exp c.f
($\sigma^2 = 100$,$\beta = 0.5$)



(H) Exp c.f
($\sigma^2 = 1$,$\beta = 5$)

FIGURE 3.2. Realizations of GRF $Z(\cdot) \sim \mathcal{GP}(\mu(\cdot), c(\cdot,\cdot))$ when $s \in [0,5]^2$ (using same seed)

4. Strong stationarity

Note 20. Assume $\mathcal{S}=\mathbb{R}^d$ for simplicity. [1]

Definition 21. A random field $Z=(Z_s)_{s\in\mathbb{R}^d}$ is strongly stationary if for all finite sets consisting of $s_1,...,s_n\in\mathbb{R}^d$, $n\in\mathbb{N}$, for all $k_1,...,k_n\in\mathbb{R}$, and for all $h\in\mathbb{R}^d$

$$\Pr\left(Z\left(s_1+h\right)\leq k_1,...,Z\left(s_n+h\right)\leq k_n\right)=\Pr\left(Z\left(s_1\right)\leq k_1,...,Z\left(s_n\right)\leq k_n\right)$$

5. Weak stationarity (or second order stationarity)

Note 22. Yuh... strong stationary may be a too “restricting” a characteristic for our modeling... Perhaps, we could only restrict the first two moments them properly; notice Def. 21 implies that, given $\mathrm{E}\left(Z_s^2\right)<\infty$, it is $\mathrm{E}\left(Z_s\right)=\mathrm{E}\left(Z_{s+h}\right)=$ contst... and $\mathrm{Cov}\left(Z_s,Z_{s'}\right)=\mathrm{Cov}\left(Z_{s+h},Z_{s'+h}\right)\overset{h=-s'}{=}\mathrm{Cov}\left(Z_{s-s'},Z_0\right)=$ funct of lag...

Definition 23. A random field $Z=(Z_s)_{s\in\mathbb{R}^d}$ is weakly stationary (or second order stationary) if, for all $s,s'\in\mathbb{R}^d$,

1. $\mathrm{E}\left(Z_s^2\right)<\infty$ (finite)
2. $\mathrm{E}\left(Z_s\right)=m$ (constant)
3. $\mathrm{Cov}\left(Z_s,Z_{s'}\right)=c\left(s'-s\right)$ for some even function $c:\mathbb{R}^d\rightarrow\mathbb{R}$ (lag dependency)

Definition 24. Weakly (or second order) stationary covariance function is called the c.f. of a weakly stationary stochastic process.

6. Covariogram

Note 25. The definition of the covariogram function requires the random field to be weakly stationary.

Definition 26. Let $Z=(Z_s)_{s\in\mathbb{R}^d}$ be a weakly stationary random field. The covariogram function of $Z=(Z_s)_{s\in\mathbb{R}^d}$ is defined by $c:\mathbb{R}^d\rightarrow\mathbb{R}$ with

$$c\left(h\right)=\mathrm{Cov}\left(Z_s,Z_{s+h}\right),\ \forall s\in\mathbb{R}^d.$$

Example 27. For the Gaussian c.f. $c\left(s,t\right)=\sigma^2\exp\left(-\beta\left\|s-t\right\|_2^2\right)$ in (Ex. 17(2)), we may denote just

$$c\left(h\right)=c\left(s,s+h\right)=\sigma^2\exp\left(-\beta\left\|h\right\|_2^2\right)\tag{6.1}$$

Observe that, in Figs 3.1 &3.2, the smaller the β, the smoother the realization (aka slower changes). One way to justify this observation is to think that smaller values of β essentially bring the points closer by re-scaling spatial lags $h$in the c.f.

[1] Otherwise, we should set $s,s'\in\mathcal{S}$, $h\in\mathcal{H}$, such as $\mathcal{H}=\left\{h\in\mathbb{R}^d:s+h\in\mathcal{S}\right\}$.

Proposition 28. *If $c : \mathbb{R}^d \to \mathbb{R}$ is the covariogram of a weakly stationary random field $Z = (Z_s)_{s\in\mathbb{R}^d}$ then:*

(1) $c(0) \geq 0$

(2) $c(h) = c(-h)$ *for all* $h \in \mathbb{R}^d$

(3) $|c(h)| \leq c(0) = Var(Z_s)$ *for all* $h \in \mathbb{R}^d$

(4) $c(\cdot)$ *is semi-positive definite; i.e. for all* $n \in \mathbb{N}$, $a \in \mathbb{R}^n$, *and* $\{s_1, ..., s_n\} \subseteq S$

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

Note 29. The following helps in the specification of cavariograms by considering properties of characteristic functions.

Theorem 30. *(Bochner's theorem) A continuous even real-valued function $c : \mathbb{R}^d \to \mathbb{R}$ is a covariance function of a weakly stationary random process if and only if it can be represented as*

$$c(h) = \int_{\mathbb{R}^d} \exp\left(i\omega^\top h\right) dF(\omega)$$

where $\mathrm{d}F(\omega)$ is a *symmetric positive finite measure on* $\mathbb{R}^d$.

- *Here, we will focus on cases of the form* $dF(\omega) = f(\omega)\,d\omega$ *where* $f(\cdot)$ *is called spectral density of* $c(\cdot)$ *i.e.*

  $$c(h) = \int_{\mathbb{R}^d} \exp\left(i\omega^\top h\right) f(\omega)\,d\omega$$

  In this case, $\lim_{h\to\infty} c(h) = 0$

Theorem 31. *If $c(\cdot)$ is integrable, $F(\cdot)$ is absolutely continuous with spectral density $f(\cdot)$ of $Z = (Z_s; s \in \mathcal{S})$ then by Fast Fourier transformation*

$$f(\omega) = \left(\frac{1}{2\pi}\right)^d \int_{\mathbb{R}^d} \exp\left(-i\omega^\top h\right) c(h)\,dh$$

Example 32. Consider the Gaussian c.f. $c(h) = \sigma^2 \exp\left(-\beta \|h\|_2^2\right)$ for $\sigma^2, \beta > 0$ and $h \in \mathbb{R}^d$. Then the spectral density from Thm 30 is

$$\begin{aligned} f(\omega) &= \left(\frac{1}{2\pi}\right)^d \int_{\mathbb{R}^d} \exp\left(-i\omega^\top h\right) \sigma^2 \exp\left(-\beta \|h\|_2^2\right) \mathrm{d}h \\ &= \sigma^2 \left(\frac{1}{2\pi}\right)^d \prod_{j=1}^d \int_{\mathbb{R}} \exp\left(-i\omega_j h_j - \beta h_j^2\right) \mathrm{d}h \\ &= \sigma^2 \left(\frac{1}{2\pi}\right)^d \prod_{j=1}^d \int_{\mathbb{R}} \exp\left(-\beta \left(h_j - \left(-i\omega / (2\beta)\right)\right)^2\right) \mathrm{d}h_j \\ &= \sigma^2 \left(\frac{1}{4\pi\beta}\right)^{d/2} \exp\left(-\|\omega\|_2^2 / (4\beta)\right) \end{aligned}$$

i.e. of a Gaussian form.

7. Intrinsic stationarity

Note 33. Getting greedier, we can further weaken the weak stationarity by considering lag dependent variance in the increments with purpose to be able to use more inclusive models; Def 23 implies that $\mathrm{Var}(Z_{s+h} - Z_s) = \mathrm{Var}(Z_{s+h}) + \mathrm{Var}(Z_s) - 2\mathrm{Cov}(Z_{s+h}, Z_s) = 2c(0) - 2c(h)$.

Definition 34. A random field $Z = (Z_s)_{s\in\mathbb{R}^d}$ is intrinsically stationary if, for all $h \in \mathbb{R}^d$, $(Z_{s+h} - Z_s)_{s\in\mathbb{R}^d}$ is weakly stationary; i.e.

1. $\mathrm{E}(Z_{s+h} - Z_s)^2 < \infty$
2. $\mathrm{E}(Z_{s+h} - Z_s) = m$ (constant)
3. $\mathrm{Var}(Z_{s+h} - Z_s) = 2\gamma(h)$ for some function $\gamma : \mathbb{R}^d \to \mathbb{R}$ (lag dependent)

Definition 35. Intrinsically stationary covariance function is called the c.f. of an intrinsically stationary stochastic process.

Example 36. The following covariance function is not weakly but intrinsically stationary

$$c(s,t) = \frac{1}{2}\left(\|s\|^{2H} + \|t\|^{2H} - \|t-s\|^{2H}\right), \ H \in (0,1)$$

because for $h \in \mathbb{R}^d$

$$c(s, s+h) = \frac{1}{2}\left(\|s\|^{2H} + \|s+h\|^{2H} - \|h\|^{2H}\right)$$

and

$$\frac{1}{2}\mathrm{Var}(Z_s - Z_{s+h}) = \frac{1}{2}\left(\mathrm{Var}(Z_s) + \mathrm{Var}(Z_{s+h}) - 2\mathrm{Cov}(Z_s, Z_{s+h})\right) = \frac{1}{2}\|h\|^{2H}$$

Note 37. The definition of the semi-variogram function requires the random field to be intrinsic stationarity; which is weaker assumption than weak stationary required by covariogram.

Definition 38. Let $Z = (Z_s)_{s\in\mathbb{R}^d}$ be intrinsically stationary. The semi-variogram of Z is defined by $\gamma : \mathbb{R}^d \to \mathbb{R}$ with

$$\gamma(h) = \frac{1}{2}\text{Var}(Z_{s+h} - Z_s), \ \ \forall s \in \mathbb{R}^d$$

Definition 39. Variogram of an intrinsically stationary random field is called the quantity $2\gamma(h)$.

Note 40. Let $Z = (Z_s)_{s\in\mathbb{R}^d}$ be weakly stationary with covariogram $c(\cdot)$. Then Z is intrinsic stationary with semi-variogram

$$\gamma(h) = c(0) - c(h), \ \ \forall h \in \mathbb{R}^d \tag{8.1}$$

Example 41. For the Gaussian covariance function (Ex. 27) the semi-variogram is

$$\gamma(h) = c(0) - c(h) = \sigma^2\left(1 - \exp\left(-\beta\|h\|_2^2\right)\right)$$

Proposition 42. *Properties of semi-variogrames. Let $Z = (Z_s)_{s\in\mathbb{R}^d}$ be an intrinsically stationary process.*

(1) It is $\gamma(h) = \gamma(-h)$, $\gamma(h) \geq 0$, and $\gamma(0) = 0$

(2) Semi-variogram is conditionally negative definite (c.n.d.): for all $a \in \mathbb{R}^n$ s.t. $\sum_{i=1}^n a_i = 0$, and for all $\forall\{s_1, ..., s_n\} \subseteq S$

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

(3) If $\gamma(h)$ is a semi-variogram, and A is a linear transformation in $\mathbb{R}^d$ then $\tilde{\gamma}(h) = \gamma(Ah)$ is a semi-variogram too.

(4) The following functions are semi-variograms

(a) $\gamma(\cdot) = \sum_{i=1}^n a_i\gamma_i(\cdot)$, if $a_i \geq 0$, and $\{\gamma_i(\cdot)\}$ are semi-variograms

(b) $\gamma(\cdot) = \int \gamma_u(\cdot)\, dF(u)$, if $\gamma_u(\cdot)$ is a semi-variogram parametrized by $u \sim F$

(c) $\gamma(\cdot) = \lim_{n\to\infty}\gamma_n(\cdot)$ if $\gamma_n(\cdot)$ is semi-variogram and the limit exists

(5) Consider intrinsically stationary stochastic processes $Y = (Y_s)_{s\in\mathbb{R}^d}$ and $E = (E_s)_{s\in\mathbb{R}^d}$ where Y and E are independent each other. Let $Z_s = Y_s + E_s$. Then

$$\gamma_Z(h) = \gamma_Y(h) + \gamma_E(h)$$

8.1. **Behavior of variogram (Nugget effect, Sill, Range).** The variogram $\gamma(h)$ is very informative when plotted against the lag h, below we discuss some of the characteristics of it, using Fig. 8.1

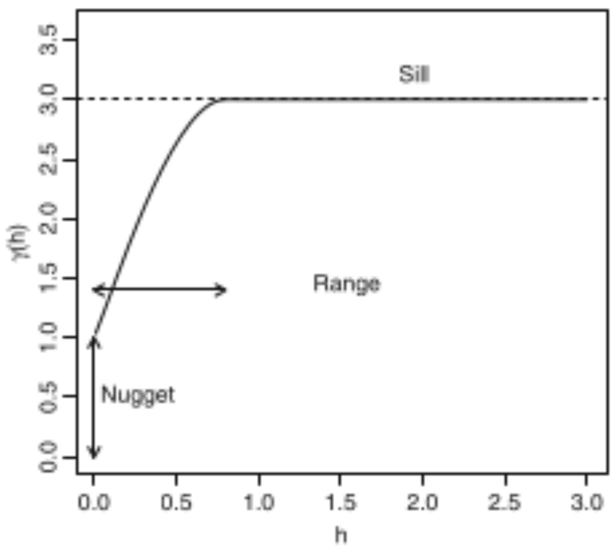


FIGURE 8.1. Variogram's characteristics

Note 43. A semivariogram tends to be an increasing function of the lag $\|h\|$. Recall in weakly stationary processes, $\gamma(h) = c(0) - c(h)$ where common logic suggests that $c(h)$ is decreases with $\|h\|$.

Note 44. If $\gamma(h)$ is a positive constant for all lags $h \neq 0$, then $Z(s_1)$ and $Z(s_2)$ are uncorrelated regardless of how close $s_1$and s_2 are; and $Z = (Z_s)_{s \in \mathbb{R}^d}$ is often called white noise.

Note 45. Conversely, a non zero slope of the variogram indicates structure.

Nugget Effect.

Note 46. Nugget effect is the semivariagram's limiting value

$$\sigma_{\varepsilon}^2 = \lim_{\|h\| \to 0} \gamma(h)$$

In particular when $\sigma_{\varepsilon}^2 \neq 0$.

Note 47. Nugget effect $\sigma_{\varepsilon}^2 \neq 0$ may expected or assumed to appear due to (1) measurement errors (e.g., if we collect repeated measurements at the same location s) or (2) due to some microscale variation causing discontinuity in the origin that cannot be detected from the data i.e. the spatial gaps because we collect a finite set of measurements at spatial locations. Hence theoretically, we could consider a more detailed decomposition $\sigma_{\varepsilon}^2 = \sigma_{\text{MS}}^2 + \sigma_{\text{MS}}^2$ where σ_{MS}^2 refers to the microscale and σ_{MS}^2 refers to the measurement error; however (my experience) this is non-identifiable.

Note 48. For a continuous processes $Z = (Z_s)_{s \in \mathbb{R}^d}$, it is expected

$$\lim_{\|h\| \to 0} \mathrm{E}\left(Z_{s+h} - Z_s\right)^2 = 0$$

which is equivalent to a continuous semivariagram $\gamma(h)$ for all h, and in particular, $\lim_{\|h\|\to 0}\gamma(h) = \gamma(0) = 0$, because $\gamma(0) = 0$. However, when modeling a real problem we may need to consider (or it may appear from the data) that $\gamma(h)$ should have a discontinuity $\lim_{\|h\|\to 0}\gamma(h) = \sigma_{\varepsilon}^2 \neq 0$.

Note 49. Nugget effect is often mathematically described by considering a decomposition ;

$$Z(s) = Y(s) + \varepsilon(s) \tag{8.2}$$

where Y can be a continuous stationary process with $\gamma_Y(\cdot)$, and ε can be a process (called errors-in-variables model) with (nugget) semivariogram $\gamma_{\varepsilon}(h) = \sigma_{\varepsilon}^2 1(h \neq 0)$. In this case,

$$\gamma_Z(h) = \gamma_Y(h) + \gamma_{\varepsilon}(h) \overset{\|h\|\to 0}{\to} \sigma_{\varepsilon}^2$$

Sill.

Definition 50. Sill is the variogram's limiting value $\lim_{\|h\|\to\infty}\gamma(h)$.

Note 51. For weakly stationary processes the sill is always finite. However, for intrinsic processes, the sill may be infinite.

Partial sill.

Definition 52. Partial sill is $\lim_{\|h\|\to\infty}\gamma(h) - \lim_{\|h\|\to 0}\gamma(h)$ which takes into account the nugget.

Range.

Definition 53. Range is the distance at which the semivariogram reaches the Sill; it can be infinite.

Other.

Note 54. An abrupt change in slope indicates the passage to a different structuration of the values in space. This is often modeled via decompoditions of processes with different semivariograms as in (8.2).

9. Isotropy

Note 55. Isotropy as a notion imposes the assumption of "rotation invariance" in the stochastic process.

Definition 56. An intrinsic stochastic process $(Z_s)_{s\in\mathbb{R}^d}$ is isotropic iff

$$\forall s,t \in \mathcal{S},\ \frac{1}{2}\text{Var}(Z_s - Z_t) = \gamma(\|t-s\|)\,,\ \text{for some function } \gamma{:}\mathbb{R}^+ \to\mathbb{R}. \tag{9.1}$$

Definition 57. Isotropic semi-variogram $\gamma : \mathbb{R}^d \to \mathbb{R}$ is the semi-variogram of the isotropic stochastic process. (sometimes for simplicity of notation we use $\gamma : \mathbb{R}^+ \to \mathbb{R}$ with $\gamma(\|h\|) = \frac{1}{2}\text{Var}(Z_s - Z_{s-h})$.

Definition 58. Isotropic covariance function $C : \mathcal{S} \times \mathcal{S} \to \mathbb{R}$ is called the covariance function satisfying (9.1).

Definition 59. Isotropic covariogram $c : \mathbb{R}^d \to \mathbb{R}$ of a weakly stationary process is the covariogram associated to an isotropic semi-variogram (sometimes for simplicity of notation we use $c : \mathbb{R}^+ \to \mathbb{R}$ with $c(\|h\|)$ from (9.1)).

9.1. **Parametric forms of frequently used isotropic covariance functions.**

Note 60. Semi-variogrames are recovered from $\gamma(h) = c(0) - c(h)$.

9.1.1. *Nugget-effect.* For $\sigma^2 > 0$,

$$c(h) = \sigma^2 1_{\{0\}}(\|h\|).$$

It is associate to white noise. It is used to model a discontinuity in the origin of the covariogram / sem-variogram.

9.1.2. *Matern c.f.* For $\sigma^2 > 0$, $\phi > 0$, and $\nu \geq 0$

$$c(h) = \sigma^2 \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\frac{\|h\|}{\phi}\right)^{\nu} K_{\nu}\left(\frac{\|h\|}{\phi}\right)$$

Parameter ν controls the variogram's regularity at 0 which in turn controls the quadratic mean (q.m.) regularity of the associated process. For $\nu = 1/2$, we get the exponential c.f.,

$$c(h) = \sigma^2 \exp\left(-\frac{1}{\phi}\|h\|_1\right)$$

which is not differentiable at $h = 0$, while for $\nu \to \infty$, we get the Gaussian c.f.

$$c(h) = \sigma^2 \exp\left(-\frac{1}{\phi}\|h\|_2^2\right)$$

which is infinite differentiable. ϕ is a range parameter, and σ^2 is the (partial) sill parameter.

9.1.3. *Spherical c.f.* For $\sigma^2 > 0$ and $\phi > 0$

$$c(h) = \begin{cases} \sigma^2\left(1 - \frac{3}{2}\frac{\|h\|_1}{\phi} + \frac{1}{2}\left(\frac{\|h\|_1}{\phi}\right)^3\right) & \|h\|_1 \leq \phi \\ \sigma^2 & \|h\|_1 > \phi \end{cases}$$

The c.f. starts from its maximum value b, then steadily decreases, and finally vanishes when its range ϕ is attained. ϕ is a range parameter, and σ^2 is the (partial) sill parameter.

10. Anisotropy

(...)

11. Estimating the semi-variogram (and covariogram)

11.1. Non-parametric estimation.

Note 61. Smoothed Matheron estimator $\hat{\gamma}(\cdot)$ of semi-variogram $\gamma(\cdot)$ is

$$\hat{\gamma}_{\mathrm{M}}(h)=\frac{1}{2\left|N_r(h)\right|}\sum_{\forall(s,t)\in N_r(h)}\left(Z(s)-Z(t)\right)^2 \tag{11.1}$$

where $N_r(h)=\left\{(s,t)\in\mathcal{S}^2:t-s\in B_r(h)\right\}$ contains all the pairs of spatial points points whole difference is in a ball $B_r(h)$ centered at h with radius $r>0$. $-\hat{\gamma}_{\mathrm{M}}$ considers all pairs of observations that are 'approximately' h apart and averages out.

Note 62. The choice of ϵ is an art, and a trade-off between variance and bias, similar to the bin length in histograms.

11.2. Parametric estimation.

Note 63. Smoothed Matheron estimator (11.1) does not necessarily satisfies semivariogram properties, such as negative definiteness. To address this we use a parametric family of appropriate semi-variogram functions and tune them against data.

Note 64. A parametric family of semivariograms is the Matern semivariogram model

$$\gamma(h)=\sigma^2\left(1-\frac{2^{1-\nu}}{\Gamma(\nu)}\left(\frac{\|h\|}{\alpha}\right)^{\nu}K_{\nu}\left(\frac{\|h\|}{\alpha}\right)\right)$$

with tunable parameters: the 'variance' $\sigma^2>0$, 'range' $\alpha>0$, shape $\nu>0$. Here, $K_\nu(\cdot)$ denotes the modified Bessel function of the second kind and order ν.

Proposition 65. *(Checking variogram's validity) A continuous function* $2\gamma(\cdot)$*with* $\gamma(0)=0$ *is a valid variogram iff*

$$\lim_{\|h\|\to\infty}\frac{\gamma(h)}{\|h\|^2}=0$$

Proposition 66. *(Checking variogram's validity) A continuous function* $2\gamma(\cdot)$*with* $\gamma(0)=0$ *is a valid variogram iff* $\exp(-a\gamma(\cdot))$ *is positive definite for any* $a>0$

Example 67. Checking the Gaussian semi-variogram in Ex 41, it is

$$\lim_{\|h\|\to\infty}\frac{\gamma(h)}{\|h\|^2}=\lim_{\|h\|\to\infty}\frac{\sigma^2\left(1-\exp\left(-\beta\|h\|_2^2\right)\right)}{\|h\|^2}=-\lim_{\|h\|\to\infty}\frac{\exp\left(-\beta\|h\|_2^2\right)}{\|h\|^2}=0.$$

Yet $\gamma(h)=\|h\|^2$ is another variogram as $\exp\left(-\beta\|h\|_2^2\right)$ is a c.f. and hence positive definite.

Note 68. Given a non-parametric estimator $\hat{\gamma}(\cdot)$ (e.g., Matheron (11.1)), and family of semi-variograms γ_θ parameterized by the unknown θ, we can learn θ by

Ordinary least squares:

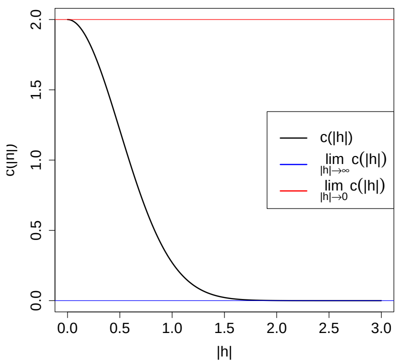
$$\hat{\theta} = \arg\min_\theta \left(\sum_j \left(\hat{\gamma}(h_j) - \gamma_\theta(h_j) \right)^2 \right)$$

Weighted least squares:

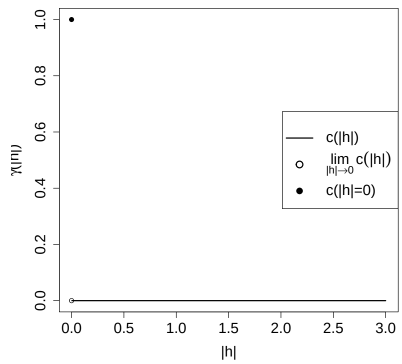
$$\hat{\theta} = \arg\min_\theta \left(\sum_j \varpi_j(\theta) \left(\hat{\gamma}(h_j) - \gamma_\theta(h_j) \right)^2 \right)$$

for some weights, for instance $\varpi_j(\theta) = |N_r(h_j)|$ or $\varpi_j(\theta) = |N_r(h_j)| / \gamma_\theta(h_j)$.

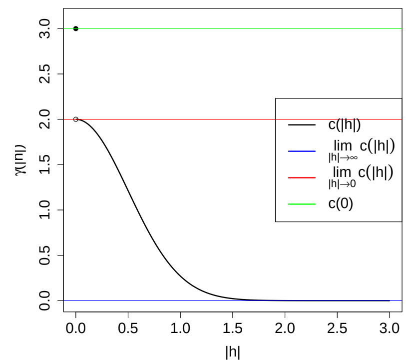
If we assume a specific distribution...



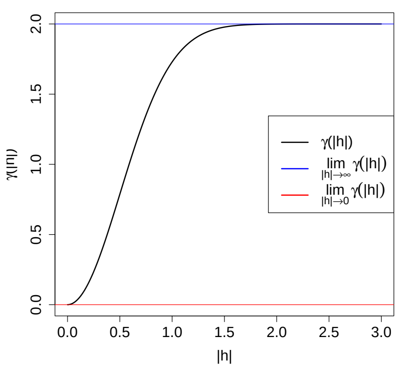
(A) $c_{\text{Gauss}}\left(\cdot\right)$



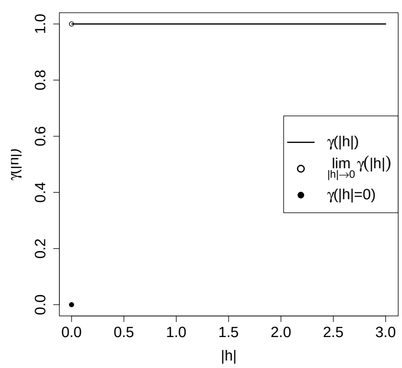
(B) $c_{\text{Nugget}}\left(\cdot\right)$



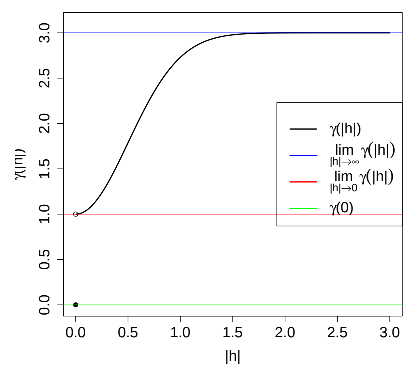
(C) $c_{\text{Gauss}}\left(\cdot\right)$ + $c_{\text{Nugget}}\left(\cdot\right)$



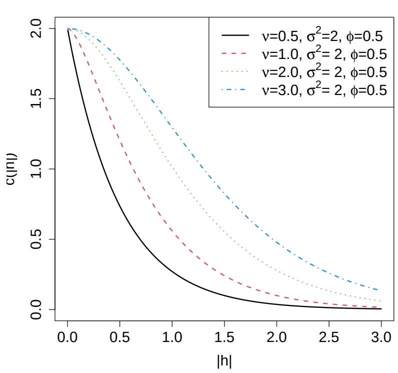
(D) $\gamma_{\text{Gauss}}\left(\cdot\right)$



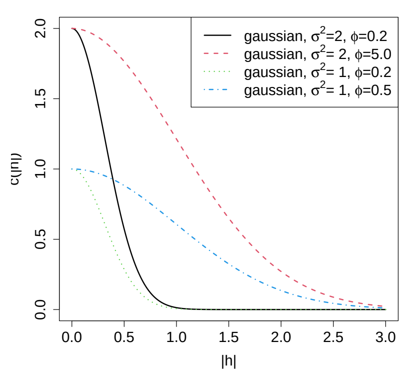
(E) $\gamma_{\text{Nugget}}\left(\cdot\right)$



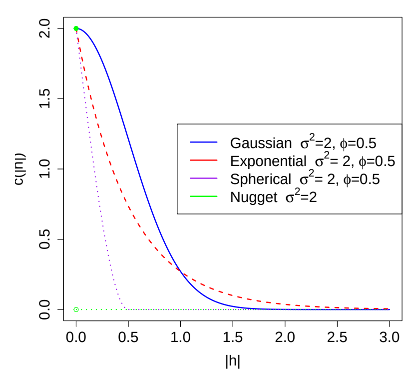
(F) $\gamma_{\text{Gauss}}\left(\cdot\right)$ + $\gamma_{\text{Nugget}}\left(\cdot\right)$



(G) c_{Matern}



(H) $c_{\text{Gauss}}\left(\cdot;\sigma^2,\phi\right)$



(I) $c_{.}\left(\cdot\right)$

FIGURE 11.1. Covariogrames $c\left(\cdot\right)$ and semivariogrames $\gamma\left(\cdot\right)$