Lecture notes part 2: Point referenced data modeling / Geostatistics

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

Aim. To introduce point referenced data modeling (geostatistics) with particular focus on concepts spatial variables, random fields, semi-variogram, kriging, change of support, multivariate geostatistics, for Bayesian and classical inference.

Reading list & references:

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

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

[3] Chiles, J. P., & Delfiner, P. (2012). Geostatistics: modeling spatial uncertainty (Vol. 713). John Wiley & Sons.

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

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

Part 1. **Basic stochastic models & related concepts for model building**

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

1. Random fields (or Stochastic processes)

Definition 2. A random field (or stochastic process, or random function) $Z = (Z(s); s \in \mathcal{S})$ 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 $(\Omega, \mathfrak{F}, \mathrm{pr})$ 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 random field is defined, and $\mathcal{Z}$ is called the state space of the field.

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

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

The family of all finite-dimensional distributions (or fidi's) of Z is called the spatial distribution of the process .

Note 4. According to Kolmogorov Theorem 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* $\mathfrak{p}$

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

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*

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

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

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

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

is a well defined random field as it satisfies Theorem 5.

1.1. **Mean and covariance functions.**

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

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

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

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

$$c(s, s') = \mathrm{Cov}(Z(s), Z(s')) = \mathrm{Cov}\left(\sum_{i=1}^{n} Z_i X_i(s), \sum_{j=1}^{n} Z_j X_j(s')\right)$$

$$= \sum_{i=1}^{n} X_i(s) \sum_{j=1}^{n} X_j(s') \underbrace{\mathrm{Cov}(Z_i, Z_j)}_{1(i = j)} = \sum_{i=1}^{n} X_i(s) X_i(s')$$

1.1.1. *Construction of covariance functions.*

Note 9. What follows provides the means for checking and constructing covariance functions.

Proposition 10. *The function $c : \mathcal{S} \times \mathcal{S} \to \mathbb{R}$, $\mathcal{S} \subseteq \mathbb{R}^d$ is a covariance function iff $c(\cdot, \cdot)$ is* semi-positive definite; i.e.

$$\forall n \in \mathbb{N} - \{0\} , \forall (a_1, ..., a_n) \in \mathbb{R}^n \text{ and } \forall (s_1, ..., s_n) \in \mathcal{S}^n : \sum_{i=1}^{n} \sum_{j=1}^{n} a_i a_j c(s_i, s_j) \geq 0$$

or in other words, the Gram matrix $(c(s_i, s_j))_{i,j=1}^{n}$ is non-negative definite for any $\{s_i\}_{i=1}^{n}$, $n \in \mathbb{N} - \{0\}$.

Example 11. $c(s, s') = 1(\{s = s'\})$ is a proper covariance function because

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

Note 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 Proposition 10, 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 random fields (or Second order processes)

Note 13. We introduce a particular class of random fields whose mean and covariance functions exist and which can be used for spatial data modeling.

Definition 14. Second order random field (or second order process) $(Z(s) ; s \in \mathcal{S})$ is called the random field where $\mathrm{E}((Z(s))^2) < \infty$ for all $s \in \mathcal{S}$.

Example 15. In second order random field $(Z(s) ; s \in \mathcal{S})$ the associated mean function $\mu(\cdot)$ and covariance function $c(\cdot, \cdot)$ exist, because $c(s, t) = \mathrm{E}(Z(s) Z(t)) - \mathrm{E}(Z(s)) \mathrm{E}(Z(t))$ for $s, t \in \mathcal{S}$.

3. Gaussian random field (or Gaussian process)

Note 16. Gaussian random field (GRF) is a particular class of second order random field which is widely used in spatial data modeling due to its computational tractability.

Also

Example

Definition 17. $(Z(s) ; s \in \mathcal{S})$ is a Gaussian random field (GRF) or Gaussian process (GP) on $\mathcal{S}$ if for any $n \in \mathbb{N}$ and for any finite set $\{s_1, ..., s_n ; s_i \in \mathcal{S}\}$, the random vector $(Z(s_1), ..., Z(s_n))^\top$
follows a multivariate normal distribution.

Proposition

Proposition 18. *A GP $(Z(s) ; s \in \mathcal{S})$ 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 19. Hence, we denote the GP as $Z(\cdot) \sim \mathcal{GP}(\mu(\cdot), c(\cdot, \cdot))$.

Note 20. When using GP for spatial modeling we just need to specify its functional parameters i.e. the mean and covariance functions.

Note 21. Popular forms of mean functions are polynomial expansions, such as $\mu(s) = \sum_{j=0}^{p-1} \beta_j s^j$ for tunable unknown parameter β. A popular form of covariance functions (c.f.), for tunable unknown parameters $\phi > 0$, and $\sigma^2 > 0$, are

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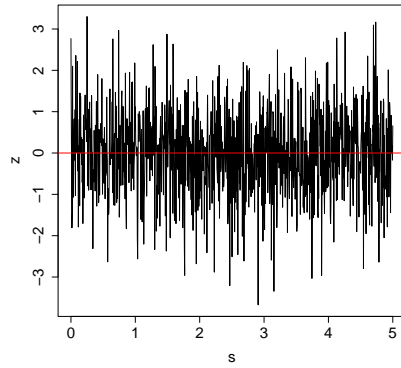(3) Nugget c.f. $c(s, s') = \sigma^2 1(s = s')$

Example 22. Recall your linear regression lessons where you specified the sampling distribution to be $y_x|\beta, \sigma^2 \overset{\text{ind}}{\sim} \text{N}(x^\top \beta, \sigma^2)$, $\forall x \in \mathbb{R}^d$. Well that can be considered as a GP $Z(\cdot) \sim \mathcal{GP}(\mu(\cdot), c(\cdot, \cdot))$ with $\mu(x) = x^\top \beta$ and $c(x, x') = \sigma^2 1(x = x')$ in (3).

Example 23. Figures 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.
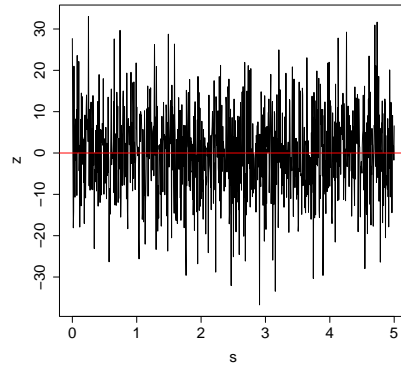
Algorithm 1 R script for simulating from a GP $(Z(s); s \in \mathbb{R}^1)$ with $\mu(s) = 0$ and $c(s,t) = \sigma^2 \exp\left(-\phi \|s - t\|_2^2\right)$

```
# set the GP parameterized mean and covariance function
mu_fun <- function(s) { return (0) }
cov_fun_gauss <- function(s,t,sig2,phi) {
    return ( sig2*exp(-phi*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 ;
phi_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,phi_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 (Figures 3.1a & 3.1b ; Figures 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 ; Figures 3.2c & 3.2d), and the spatial dependence / frequency of the ups and downs is controlled by β (Figures 3.1d & 3.1e ; Figures 3.2d & 3.2e). Realizations with different c.f. have different behavior (Figures 3.1a, 3.1d & 3.1e ; Figures 3.2a, 3.2d & 3.2e)

Created on 2024/11/04 at 11:00:38   by Georgios Karagiannis

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

(B) Nugget c.f
($\sigma^2 = 100$)
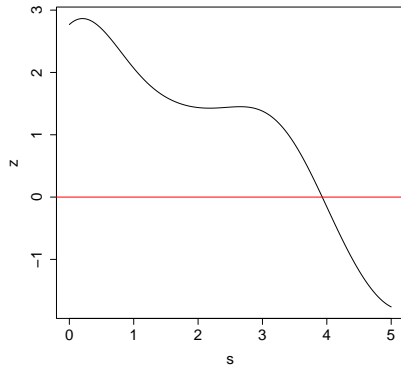
(C) Gauss c.f
($\sigma^2 = 1, \phi = 0.5$)

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

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

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

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

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

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

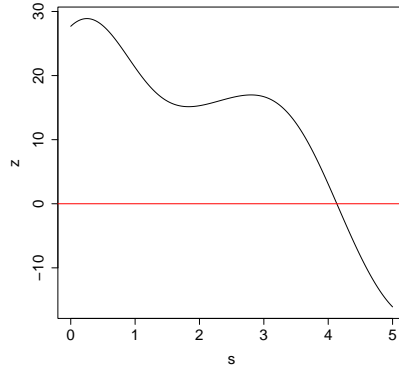Created on 2024/11/04 at 11:00:38   by Georgios Karagiannis
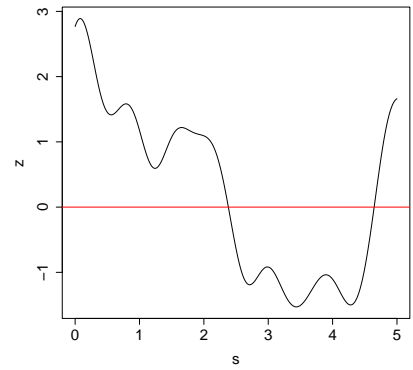
(A) Nugget c.f
$(\sigma^2 = 1)$

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

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

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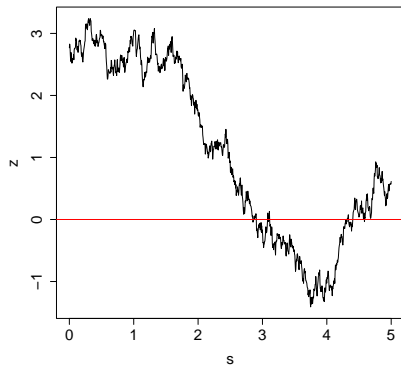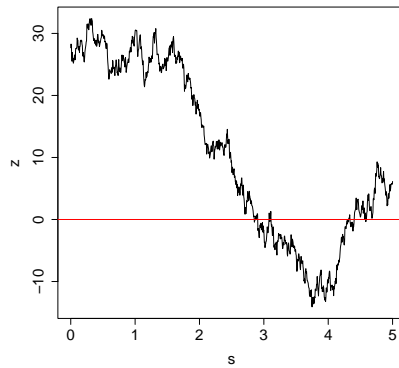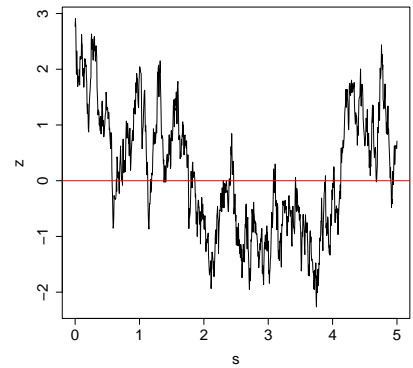$(\sigma^2 = 100, \phi = 0.5)$

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

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

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

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

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)
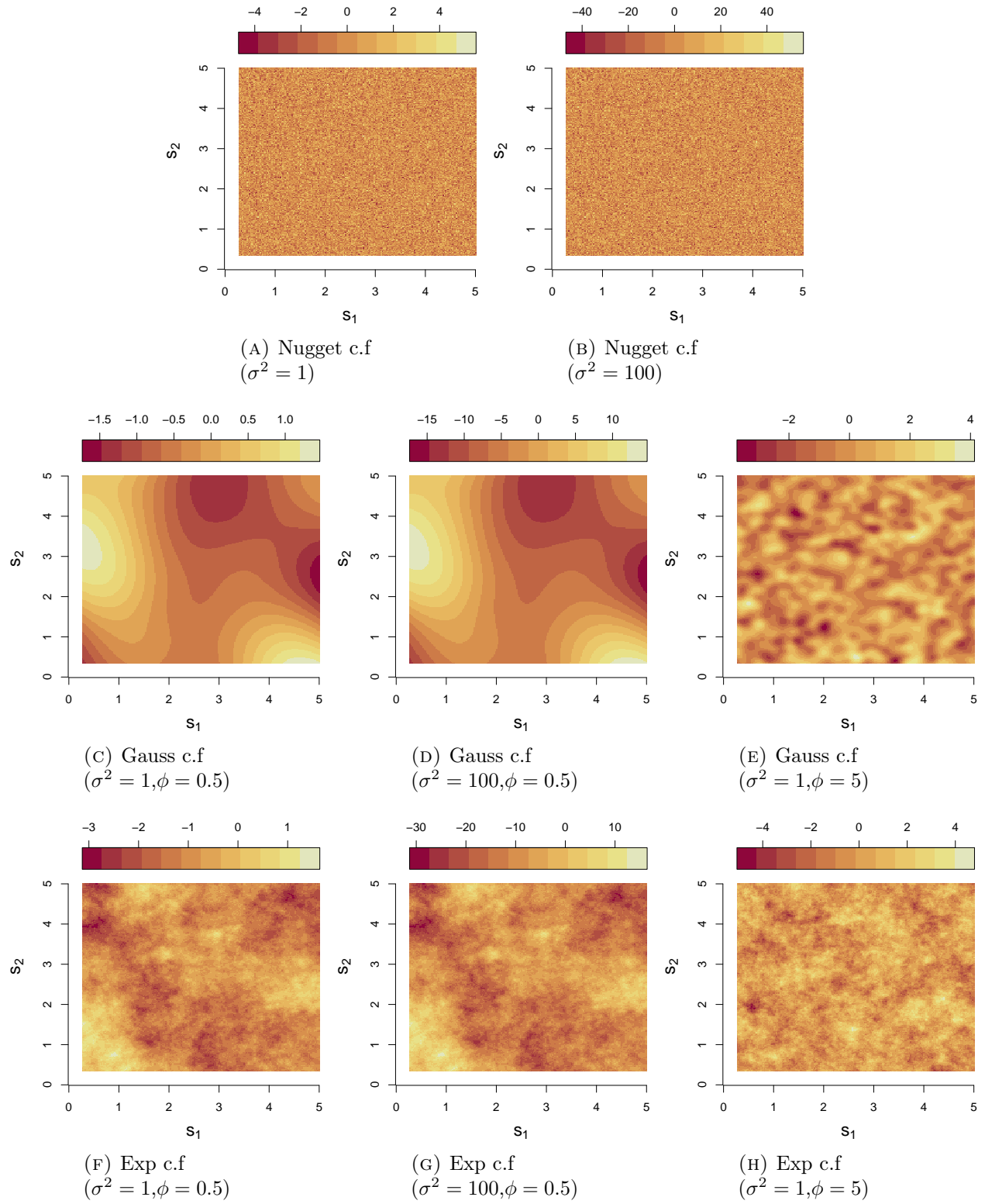
Created on 2024/11/04 at 11:00:38   by Georgios Karagiannis

4. Strong stationarity

Note 24. We introduce a specific behavior of random field to build our models.

Notation 25. Formally, we define the separation (or lag) set as $\mathcal{H} = \left\{ h \in \mathbb{R}^d : s \in \mathcal{S}, s + h \in \mathcal{S} \right\}$ where $\mathcal{S} \subseteq \mathbb{R}^d$ is the spatial domain for $d = 1, 2, 3, \dots$. However, we will consider cases where $\mathcal{S} = \mathbb{R}^d$ and $\mathcal{H} = \mathbb{R}^d$ for $d = 1, 2, 3, \dots$ in Euclidean spaces.

Definition 26. A random field $(Z(s); s \in \mathcal{S})$ is strongly stationary on $\mathcal{S}$ if for all finite sets consisting of $s_1, \dots, s_n \in \mathcal{S}$, $n \in \mathbb{N}$, for all $k_1, \dots, k_n \in \mathbb{R}$, and for all $h \in \mathcal{H}$

$$\mathrm{pr}\left(Z(s_1 + h) \leq k_1, \dots, Z(s_n + h) \leq k_n \right) = \mathrm{pr}\left(Z(s_1) \leq k_1, \dots, Z(s_n) \leq k_n \right)$$

Note 27. Yuh... strong stationary may represent a behavior being too "restrictive" to be used for spatial data modeling as it is able to represent only limiting number of spatial dependencies.

5. Weak stationarity (or second order stationarity)

Note 28. We introduce another weaker behavior of random field able to represent a larger class of spatial dependencies.

Note 29. Instead of working with the "restrictive" strong stationarity, we could just properly specify the behavior of the first two moments only; notice that Definition 26 implies that, given $\mathrm{E}\left((Z(s))^2 \right) < \infty$, it is $\mathrm{E}(Z(s)) = \mathrm{E}(Z(s+h)) = \mathrm{contst}\dots$ and $\mathrm{Cov}(Z(s), Z(s')) = \mathrm{Cov}(Z(s+h), Z(s'+h)) \overset{h=-s'}{=} \mathrm{Cov}(Z(s-s'), Z(0)) = \mathrm{funct}$ of lag...

Definition 30. A random field $(Z(s); s \in \mathcal{S})$ is called stationary random field (s.r.f.) (or weakly stationary or second order stationary) if it has constant mean and translation invariant covariance; i.e. for all $s, s' \in \mathbb{R}^d$,

(1) $\mathrm{E}\left((Z(s))^2 \right) < \infty$ (finite)
(2) $\mathrm{E}(Z(s)) = \mu$ (constant)
(3) $\mathrm{Cov}(Z(s), Z(s')) = c(s' - s)$ for some even function $c : \mathcal{S} \to \mathbb{R}$ (lag dependency)

Definition 31. Stationary (or weakly or second order stationary) covariance function is called the c.f. of a stationary random field.

6. Covariogram

Note 32. We introduce the covariogram function able to express many aspects of the behavior of a (weakly) stationary random field and hence be used as statistical descriptive tool.

Created on 2024/11/04 at 11:00:38 by Georgios Karagiannis

Definition 33. The covariogram function of a weakly stationary random field $(Z(s); s \in \mathcal{S})$ is defined by $c : \mathcal{H} \to \mathbb{R}$ with

$$c(h) = \text{Cov}(Z(s), Z(s+h)), \ \forall s \in \mathcal{S}, \ \forall h \in \mathcal{H}.$$

Example 34. For the Gaussian c.f. $c(s,t) = \sigma^2 \exp(-\phi \|s-t\|_2^2)$ in (Ex. 20(2)), we may denote just

(6.1) $$c(h) = c(s, s+h) = \sigma^2 \exp(-\phi \|h\|_2^2)$$

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

Proposition 35. *If* $c : \mathcal{H} \to \mathbb{R}$ *is the covariogram of a weakly stationary random field* $(Z(s); s \in \mathcal{S})$ *then:*

(1) $c(h) = c(-h)$ *for all* $h \in \mathcal{H}$
(2) $|c(h)| \leq c(0) = Var(Z(s))$ *for all* $h \in \mathcal{H}$
(3) $c(\cdot)$ *is semi-positive definite; i.e. for all* $n \in \mathbb{N} - \{0\}$, $(a_1, ..., a_n) \in \mathbb{R}^n$, *and* $(s_1, ..., s_n) \subseteq \mathcal{S}^n$

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

Note 36. Given there is some knowledge of the characteristic functions of a suitable distribution, the following spectral representation theorem helps in the specification of a suitable covariogram.

Theorem 37. *(Bochner's theorem) Let* $c : \mathbb{R}^d \to \mathbb{R}$ *be a continuous even real-valued function for* $d \geq 1$. *Then* $c(\cdot)$ *is positive semidefinete (hence a covariogram of a stationary random field) if and only if it can be represented as*

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

where F *is a symmetric positive finite measure on* $\mathbb{R}^d$ *called spectral measure.*

Note 38. In our course, we focus on cases where F has a density $f(\cdot)$ i.e. $dF(\omega) = f(\omega) d\omega$. $f(\cdot)$ is called spectral density of $c(\cdot)$, it is

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

and it dies as $\lim_{|h| \to \infty} c(h) = 0$

Theorem 39. *If $c(\cdot)$ is integrable, the spectral density $f(\cdot)$ can be computed by inverse 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 40. Consider the Gaussian c.f. $c(h) = \sigma^2 \exp\left(-\phi \|h\|_2^2\right)$ for $\sigma^2, \beta > 0$ and $h \in \mathbb{R}^d$. Then, by using Theorem 37, the spectral density is

$$f(\omega) = \left(\frac{1}{2\pi}\right)^d \int_{\mathbb{R}^d} \exp\left(-i\omega^\top h\right) \sigma^2 \exp\left(-\phi \|h\|_2^2\right) \mathrm{d}h$$

$$= \sigma^2 \left(\frac{1}{2\pi}\right)^d \prod_{j=1}^{d} \int_{\mathbb{R}} \exp\left(-i\omega_j h_j - \phi h_j^2\right) \mathrm{d}h$$

$$= \sigma^2 \left(\frac{1}{2\pi}\right)^d \prod_{j=1}^{d} \int_{\mathbb{R}} \exp\left(-\phi\left(h_j - (-i\omega/(2\phi))\right)^2\right) \exp\left(-\omega_j^2/(4\phi)\right) \mathrm{d}h_j$$

$$= \sigma^2 \left(\frac{1}{4\pi\phi}\right)^{d/2} \exp\left(-\|\omega\|_2^2/(4\phi)\right)$$

i.e. it has a Gaussian form.

Definition 41. Let $(Z(s) : s \in \mathcal{S})$ be a weakly stationary random field with covariogram function $c : \mathcal{H} \to \mathbb{R}$ and $c(h) = \mathrm{Cov}(Z(s), Z(s+h))$. The correlogram function $\rho : \mathcal{H} \to [-1, 1]$ is defined as

$$\rho(h) = \frac{c(h)}{c(0)}.$$

7. Intrinsic stationarity (of order zero)

Note 42. The class of (weakly) stationary random fields may not be sufficiently general enough to model a large number of important applications. E.g., in certain applications, it has been noticed that the "underline process" we wish to model presents increments whose variance

$$\mathrm{Var}(Z(s+h) - Z(s)) = \mathrm{Var}(Z(s+h)) + \mathrm{Var}(Z(s)) - 2\mathrm{Cov}(Z(s+h), Z(s))$$

increases indefinitely with $|h|$; this "process" cannot be modeled within the class of (weakly) stationary random fields whose increments have bounded variance $\mathrm{Var}(Z(s+h) - Z(s)) = 2(c(0) - c(h)) < 2c(0)$. Intrinsic stationary is a weaker assumption extending the class of models we can use.

Definition 43. A random field $(Z(s) : s \in \mathcal{S})$ is called intrinsic random field (i.r.f.) (or intrinsic stationary r.f.) if, for all $h \in \mathcal{H}$, ,

(1) $E(Z(s+h) - Z(s))^2 < \infty$

(2) $E(Z(s+h) - Z(s)) = \mu(h)$ for some function $\mu: \mathcal{H} \to \mathbb{R}$ (lag dependent)

(3) $\mathrm{Var}(Z(s+h) - Z(s)) = 2\gamma(h)$ for some function $\gamma: \mathcal{H} \to \mathbb{R}$ (lag dependent)

Example 44. The random field with covariance function

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

is not stationary r.f. because

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

for $h \in \mathcal{H}$ but it intrinsic r.f. because

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

Example 45. For an i.r.f. $(Z(s): s \in \mathcal{S})$ with $\mu(h) = 0$, it can be shown that

$$(7.1) \qquad \mathrm{Cov}(Z(t) - Z(s), Z(v) - Z(u)) = \gamma(t-u) + \gamma(s-v) - \gamma(s-u) - \gamma(t-u)$$

by taking expectations from

$$2(a-b)(c-e) = (a-e)^2 + (a-b)^2 - (b-c)^2 - (a-c)^2$$

Note 46. The price to be paid for i.r.f. offering a larger class of models by setting the assumptions on the increments only, is involve an indeterminacy regarding the actual r.f. $Z(s)$; E.g. i.r.f. . $(Z(s): s \in \mathcal{S})$ and $(Z(s) + U: s \in \mathcal{S})$ where U a single variable leave (2) and (3) in Def 43 unchanged. When this causes problems, usual trick are: (a) "registration" (Example 47), i.e. consider an additional non-used specific site $s_0 \in \mathcal{S}$ (called origin) at which a value is known $Z(s_0) = z_0$ and try to work out (b) impose restrictions int eh increments.

Example 47. To specify the moments of an i.r.f. $Z(s)$. Consider an origin $s_0 \in \mathcal{S}$ with known $Z(s_0) = z_0$, and specify the "registered" r.f. $\tilde{Z}(s) = z_0 + (Z(s) - Z(s_0))$. Then $E\left(\tilde{Z}(s)\right) = z_0 + \mu(s - s_0)$ and $\mathrm{Cov}\left(\tilde{Z}(s), \tilde{Z}(t)\right)$ computed from (7.1).

Example 48. Only the covariance of allowed linear combinations can be represented w.r.t. $\gamma(\cdot)$. I.e.

$$\mathrm{Cov}\left(\sum_{i=1}^{n} a_i Z(s_i), \sum_{j=1}^{n} b_j Z(s_j)\right) = \sum_{i=1}^{n} a_i \sum_{j=1}^{n} b_j \mathrm{Cov}(Z(s_i), Z(s_j))$$

Created on 2024/11/04 at 11:00:38 by Georgios Karagiannis

assuming Z is i.r.f. hence covariogram may not be defined, we consider origin $s_0 \in \mathcal{S}$ with $Z(s_0)$ and we restrict to the sum-to-zero linear combinations. Hence, by (7.1)

$$
\begin{aligned}
\text{Cov}\left(\sum_{i=1}^{n} a_i Z(s_i), \sum_{j=1}^{n} b_j Z(s_j)\right) &= \sum_{i=1}^{n} a_i \sum_{j=1}^{n} b_j \text{Cov}(Z(s_i) - Z(s_0), Z(s_j) - Z(s_0)) \\
&= \sum_{i=1}^{n} a_i \sum_{j=1}^{n} b_j (\gamma(s_i - s_0) + \gamma(s_j - s_0) - \gamma(s_j - s_i)) \\
&= -\sum_{i=1}^{n} \sum_{j=1}^{n} a_i b_j \gamma(s_j - s_i)
\end{aligned}
$$

8. Incremental mean function

Definition 49. Incremental mean function (or drift) of the intrinsic random field $(Z(s) : s \in \mathcal{S})$ is defined as $\mu : \mathcal{H} \to \mathbb{R}$ with $\mu(h) = \mathrm{E}(Z(s+h) - Z(s))$.

Example 50. Let $(Z(s) : s \in \mathcal{S})$ be an intrinsic random field, the incremental drift is linear

$$
\mu(h) = h^\top \beta \tag{8.1}
$$

for some $\beta \in \mathbb{R}^d$. Indeed, it is

$$
\begin{aligned}
\mu(h + h') = \mathrm{E}(Z(s + h + h') - Z(s)) &= \mathrm{E}(Z(s+h) - Z(s)) + \mathrm{E}(Z(s + h + h') - Z(s+h)) \\
&= \mu(h) + \mu(h'), \ \forall h, h'.
\end{aligned}
$$

Since, $\mu(\cdot)$ is continuous and $\mu(0) = 0$, than $\mu(h)$ is linear wrt h.

9. Semivariogram

Note 51. A very informative tool about the behavior of the intrinsic random field is the semivariogram function defined below.

Definition 52. The semivariogram of an intrinsic random field $(Z(s) : s \in \mathcal{S})$ is defined as $\gamma : \mathcal{H} \to \mathbb{R}$, with

$$
\gamma(h) = \frac{1}{2}\text{Var}(Z(s+h) - Z(s))
$$

Definition 53. Variogram of an intrinsic random field $(Z(s) : s \in \mathcal{S})$ is called the quantity $2\gamma(h)$.

Note 54. A stationary random field with covariogram $c(\cdot)$ and mean μ is intrinsic stationary as well with semivariogram

$$
\gamma(h) = c(0) - c(h), \tag{9.1}
$$

and constant incremental mean $\mu(h) = \mu$.

Example 55. For the Gaussian covariance function (Ex. 34) the semivariogram is

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

Proposition 56. *Properties of semivariogram. Let $(Z(s) : s \in \mathcal{S})$ be an intrinsic random field, then*

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

(2) Semivariogram is conditionally negative definite (c.n.d.): if for all $n \in \mathbb{N}$, $(a_1, ..., a_n) \subseteq \mathbb{R}^n$ s.t. $\sum_{i=1}^{n} a_i = 0$, and for all $(s_1, ..., s_n) \subseteq S^n$, it is

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

10. Behavior of semivariogram of intrinsic random fields

Note 57. The semivariogram $\gamma(h)$ is very informative when plotted against the lag h. Below we discuss some of the characteristics of it, using Figure 10.1
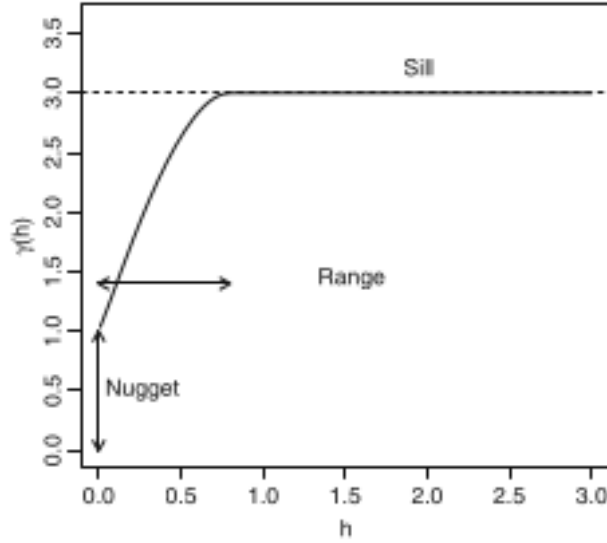
Figure 10.1. Semi Variogram's characteristics

Note 58. A semivariogram tends to be an increasing function of the lag $\|h\|$. Recall that for weakly stationary random fields with c.f. $c(\cdot)$, it is $\gamma(h) = c(0) - c(h)$ where common logic suggests that $c(h)$ is decreases with $\|h\|$.

Note 59. If $\gamma(h)$ is a positive constant for all non-zero lags $h \neq 0$, then $Z(s_1)$ and $Z(s_2)$ are uncorrelated regardless of how close s_1 and s_2 are. Then $Z(\cdot)$ is called white noise.

Note 60. Conversely, a non zero slope of the variogram indicates some structure.

Created on 2024/11/04 at 11:00:38 by Georgios Karagiannis

Nugget Effect.

Note 61. Nugget effect is the semivariagram limiting value

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

when $\sigma_\varepsilon^2 \neq 0$.

Note 62. When used for modeling, nugget effect $\sigma_\varepsilon^2 \neq 0$ may expresses (1) measurement errors (e.g., if we collect repeated measurements at the same location s) or (2) 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. Ideally, a more detailed decomposition $\sigma_\varepsilon^2 = \sigma_{\mathrm{MS}}^2 + \sigma_{\mathrm{MS}}^2$ can be considered where σ_{MS}^2 refers to the microscale and σ_{MS}^2 refers to the measurement error. However this may lead to non-identifiability, without any obvious tweak to address it.

Sill.

Definition 63. Sill is the semivariagram limiting value $\lim_{\|h\|\to\infty} \gamma(h)$.

Note 64. For intrinsic processes, the sill may be infinite or finite. For weakly random field, the sill is always finite.

Partial sill .

Definition 65. Partial sill is $\lim_{\|h\|\to\infty} \gamma(h) - \lim_{\|h\|\to 0} \gamma(h)$ which takes into account the nugget.

Range .

Note 66. Range is the distance at which the semivariogram reaches the Sill. It can be infinite or finite.

Other.

Note 67. An abrupt change in slope indicates the passage to a different structuration of the values in space. This is often modeled via decomposition of processes with different semivariograms. E.g., let independent random fields $Y(\cdot)$ and $X(\cdot)$ with different semivariograms γ_Y and γ_X,then random field $Z(\cdot)$ with $Z(s) = Y(s) + X(s)$ has semivariogram $\gamma_Z(h) = \gamma_Y(h) + \gamma_X(h)$ which may present such a behavior.

11. Isotropy

Note 68. Isotropy introduces the assumption of "rotation invariance".

Note 69. Isotropy applies to both intrinsic and (weakly) stationary random fields.

Definition 70. An intrinsic random field $(Z(s) : s \in \mathcal{S})$ is isotropic iff

$$(11.1) \qquad \forall s, t \in \mathcal{S}, \; \frac{1}{2}\mathrm{Var}\left(Z(s) - Z(t)\right) = \gamma\left(\|t - s\|\right), \; \text{ for some function } \gamma{:}\mathbb{R}_{\geq 0} \to \mathbb{R}.$$

Definition 71. Isotropic semivariogram $\gamma : \mathcal{H} \to \mathbb{R}$ is the semivariogram of the isotropic random field (sometimes for simplicity of notation we use $\gamma : \mathbb{R}_{\geq 0} \to \mathbb{R}$ with $\gamma\left(\|h\|\right) = \frac{1}{2}\mathrm{Var}\left(Z(s) - Z(s - h)\right)$.

Definition 72. Isotropic covariance function $C : \mathcal{S} \times \mathcal{S} \to \mathbb{R}$ is called the covariance function satisfying (11.1).

Definition 73. Isotropic covariogram $c : \mathcal{H} \to \mathbb{R}$ of a weakly stationary process is the covariogram associated to an isotropic semivariogram. Sometimes for simplicity of notation we use $c : \mathbb{R}_{\geq 0} \to \mathbb{R}$ with $c\left(\|h\|\right)$ from (11.1).

11.1. **Popular isotropic covariance functions.**

Note 74. Isotropic semivariograms can be computed from $\gamma(h) = c(0) - c(h)$ given covariogram $c(\cdot)$ for any h.

11.1.1. *Nugget-effect.*

Note 75. Nugget-effect covariogram takes the form

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

for $\sigma^2 > 0$. It is associate to white noise. It is used to model a discontinuity in the origin of the covariogram / sem-variogram.

11.1.2. *Matern c.f.*

Note 76. Matern covariogram takes the form

No need to memorize (11.2)

$$(11.2) \qquad\qquad c(h) = \sigma^2 \frac{2^{1-\nu}}{\Gamma(\nu)}\left(\frac{\|h\|}{\phi}\right)^{\nu} K_{\nu}\left(\frac{\|h\|}{\phi}\right)$$

for $\sigma^2 > 0$, $\phi > 0$, and $\nu \geq 0$. 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.

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

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

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

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

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

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

(G) c_{Matern}

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

(I) $c_{\cdot}(\cdot)$
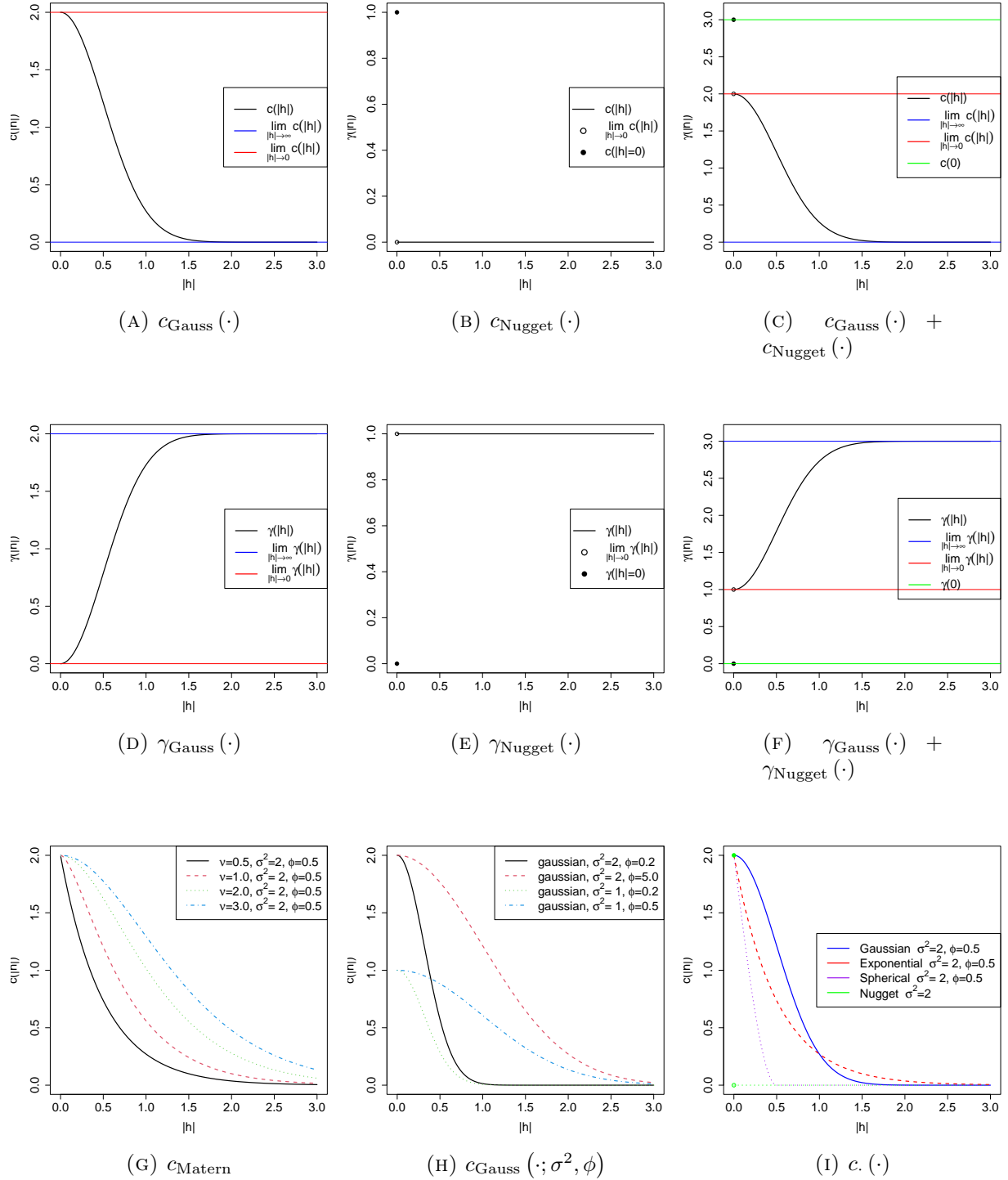
FIGURE 11.1. Covariogrames $c(\cdot)$ and semivariogrames $\gamma(\cdot)$

11.1.3. *Spherical c.f.*

Created on 2024/11/04 at 11:00:38
by Georgios Karagiannis

Note 77. [1]Spherical covariograme takes the form

$$(11.3) \qquad c(h) = \begin{cases} \sigma^2 \left(1 - \frac{3}{2} \frac{\|h\|_1}{\phi} + \frac{1}{2} \left(\frac{\|h\|_1}{\phi} \right)^3 \right) & \|h\|_1 \leq \phi \\ 0 & \|h\|_1 > \phi \end{cases}, \quad h \in \mathbb{R}^3.$$

for $\sigma^2 > 0$ and $\phi > 0$. The c.f. starts from its maximum value σ^2 at the origin, then steadily decreases, and finally vanishes when its range ϕ is reached. ϕ is a range parameter, and σ^2 is the (partial) sill parameter.

12. ANISOTROPY

Note 78. Dependence between $Z(s)$ and $Z(s+h)$ is a function of both the magnitude and the direction of separation h. This can be caused by the underlying physical process evolving differently in space (e.g., vertical and horizontal axes).

Definition 79. The semivariogram $\gamma : \mathcal{H} \to \mathbb{R}$ is anisotropic if there are h_1 and h_2 with same length $\|h_1\| = \|h_2\|$ but different direction $h_1/\|h_1\| \neq h_2/\|h_2\|$ that produce different semivariograms $\gamma(h_1) \neq \gamma(h_2)$.

Definition 80. The intrinsically random field $(Z(s) : s \in \mathcal{S})$ is anisotropic if its semivariogram is anisotropic.

Definition 81. The covariogram $c : \mathcal{H} \to \mathbb{R}$ is anisotropic if there are h_1 and h_2 with same length $\|h_1\| = \|h_2\|$ but different direction $h_1/\|h_1\| \neq h_2/\|h_2\|$ that produce different covariogram $c(h_1) \neq c(h_2)$.

Definition 82. The (weakly) stationary random field $(Z(s) : s \in \mathcal{S})$ is anisotropic if its covariogram is anisotropic.

Note 83. For brevity, below we discuss about intrinsic random fields and semivariograms, however the concepts/definitions apply to stationary random fields and covariograms when defined, as in Defs 79 & 81.

12.1. **Geometric anisotropy.**

Definition 84. The semivariogram $\gamma_{\text{g.a.}} : \mathcal{H} \to \mathbb{R}$ exhibits geometric anisotropy if it results from an A-linear deformation of an isotropic semivariogram with function $\gamma_{\text{iso}}(\cdot)$; i.e.

$$\gamma_{\text{g.a.}}(h) = \gamma_{\text{iso}}(\|Ah\|_2)$$

Note 85. Such semivariograms have the same sill in all directions but with ranges that vary depending on the direction. See Figure 12.1a.

---

[1]For it's derivation see Ch 8 in [4]
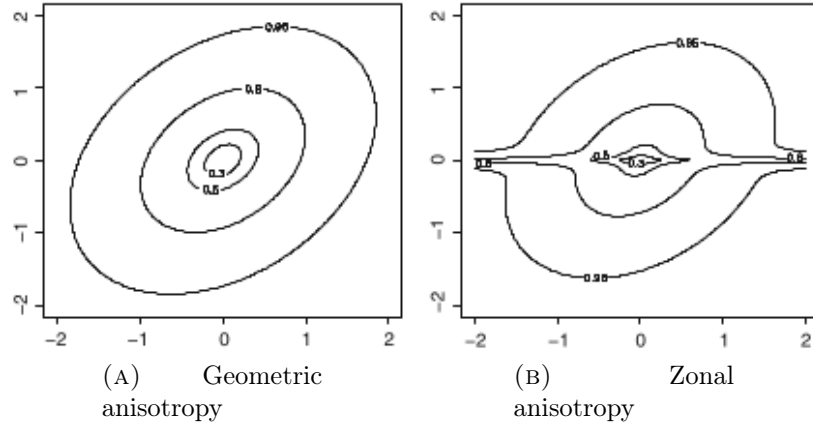
FIGURE 12.1. Isotropy vs Anisotropy
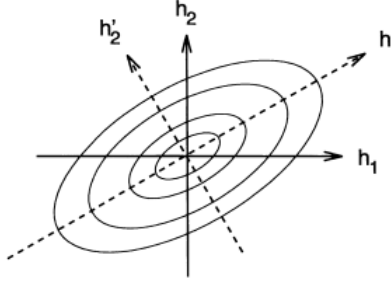
FIGURE 12.2. Rotation of the 2D coordinate system

Example 86. For instance, if $\gamma_{\text{g.a.}}(h) = \gamma_{\text{iso}}\left(\sqrt{h^\top Q h}\right)$, where $Q = A^\top A$.

Example 87. [Rotating and dilating an ellipsoid in 2D] Consider a coordinate system for $h = (h_1, ..., h_n)^\top$. We wish to find a new coordinate system for h in which the iso-semivariogram lines are spherical.

(1) [Rotate] Apply rotation matrix R to h such as $h' = Rh$. In 2D, it is

$$R = \begin{bmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{bmatrix}, \quad \text{for } \theta \in (0, 2\pi), \text{ is the rotation angle.}$$

(2) [Dilate] Apply a dilation of the principal axes of the ellipsoid using a diagonal matrix $\sqrt{\Lambda} = \text{diag}\left(\sqrt{\lambda_1}, ..., \sqrt{\lambda_n}\right)$, as $\tilde{h} = \sqrt{\Lambda} h'$.

Now the ellipsoids become spheres with radius $r = \left\|\tilde{h}\right\|_2 = \sqrt{\tilde{h}^\top \tilde{h}}$. This yields the equation of an ellipsoid in the h coordinate system

$$h^\top \left(R^\top \Lambda R\right) h = r^2$$

Created on 2024/11/04 at 11:00:38   by Georgios Karagiannis

where the diameters d_j (principal axes) of the ellipsoid along the principal directions are

$$d_j = 2r/\sqrt{\lambda_j}$$

and the principal direction is the j-th column of the rotation matrix $R_{:,j}$.

Hence the anisotropic semivariogram is $\gamma_{\text{g.a.}}(h) = \gamma_{\text{iso}}\left(\sqrt{h^\top Q h}\right)$ with $Q = R^\top \Lambda R$. This derivation extends to d dimensions.

12.2. Zonal (or stratified) anisotropy.

Definition 88. Support anisotropy is called the type of anysotropy that the semivariogram $\gamma(h)$ depends only on certain coordinates of h.

Example 89. If it is $\gamma(h = (h_1, h_2)) = \gamma(h_1)$, then I ve support anisotropy

Definition 90. Zonal anisotropy occurs when the semivariogram $\gamma(h)$ is the sum of several components each with a support anisotropy.

Example 91. Let γ' and γ'' be semivariogram with sills v' and v'' correspondingly. If it is $\gamma(h = (h_1, h_2)) = \gamma'(\|h_1\|) + \gamma''\left(\sqrt{\|h_1\| + \|h_2\|}\right)$, then I 've Zonal anisotropy because γ has a sill $v' + v''$ is direction $(0, 1)$ and a sill v' in direction $(1, 0)$.

Note 92. We have Zonal anisotropy then the semivariogram calculated in different directions suggest a different value for the sill (and possibly the range).

Note 93. If in 2D case, the sill in h_1 is larger than that in h_2, we can model zonal anysotropy of random field $(Z(s) : s \in \mathcal{S})$ by assuming $Z(s) = I(s) + A(s)$, where $I(s)$ is an isotropic random field with isotropic semivariogram γ_I along dimension of h_1 and $A(s)$ is an process with anisotropic semivariogram γ_I without effect on dimension h_1; i.e. $\gamma_Z(h) = \gamma_I(h) + \gamma_A(h)$.

12.3. Non-linear deformations.

Note 94. A (rather too general) non-stationary non-intrinsic random field model can be specified by considering semivariogram $2\text{Var}(Z(s) - Z(t)) = 2\gamma_o(\|G(s) - G(t)\|)$ such that a bijective non-linear (function) deformation $G(\cdot)$ of space $\mathcal{S}$ has been applied on the isotropic semivariogram γ_o. For instance, $\gamma_o(h) = \sigma^2 \exp(-\|h\|/\phi)$ and $G(s) = s^2$ as a deterministic function. Now, if function $G(\cdot)$ is considered as unknown, one can model it as a random field $(G(s) : s \in \mathcal{S})$ with semivariogram $2\text{Var}(G(s) - G(t)) = 2\gamma_o'(\|G'(s) - G'(t)\|)$ and so on...; then we will be talking about deep learning.

13. GEOMETRICAL PROPERTIES OF RANDOM FIELDS

Note 95. We discuss basic geometric properties of random field we will use for modeling, as it can give us a deeper intuition on how to design appropriate spatial statistical models.

Definition 96. (Continuity in quadratic mean (q.m.)) Second-order random field $(Z(s) : s \in \mathcal{S})$ is q.m. continuous at $s \in \mathcal{S}$ if

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

Note 97. Consider random field $(Z(s) : s \in \mathcal{S})$ .Then

$$\mathrm{E}\left(Z(s+h) - Z(s)\right)^2 = \left(\mathrm{E}\left(Z(s+h)\right) - \mathrm{E}\left(Z(s)\right)\right)^2 + \mathrm{Var}\left(Z(s+h) - Z(s)\right)$$

- If Z is intrinsic r.f., then

$$\mathrm{E}\left(Z(s+h) - Z(s)\right)^2 = \frac{1}{2}\gamma(h)$$

  and hence Z is q.m. continuous iff $\lim_{\|h\| \to 0} \gamma(h) = \gamma(0)$.
- If Z is stationary r.f., then

$$\mathrm{E}\left(Z(s+h) - Z(s)\right)^2 = \frac{1}{2}\left(c(0) - c(h)\right)$$

  and hence Z is q.m. continuous iff $\lim_{\|h\| \to 0} c(h) = c(0)$ (i.e. ,c is continuous).

Definition 98. Differentiable in quadratic mean (q.m.)) Second-order random field $(Z(s) : s \in \mathcal{S})$ is q.m. differentiable at $s \in \mathcal{S}$ there exist

$$(13.1) \qquad \dot{Z}(s) = \lim_{h \to 0} \frac{Z(s+h) - Z(s)}{h}. \text{ in q.m.}$$

Proposition 99. *Let $c(s,t)$ be the covariance function of $Z = (Z(s) : s \in \mathcal{S})$. Then Z is everywhere differentiable if $\frac{\partial^2}{\partial s \partial t} c(s,t)$ exists and it is finite. Also, $\frac{\partial^2}{\partial s \partial t} c(s,t)$ is the covariance function of (13.1).*

Example 100. The process with exponential c.f. $c(h) = \sigma^2 \exp\left(-|h|/\phi\right)$ is continuous because $\lim_{h \to 0} c(h) = \sigma^2 = c(0)$ but not differentiable because $\frac{\partial^2}{\partial h^2} c(h)$ does not exist at $h = 0$.

Part 2. **Model building & related parametric inference**

14. The Geostatistical model (the big picture)

14.1. **Linear Model of Regionalization.**

Note 101. A spatial phenomenon can be thought as being the sum of several independent subphenomena acting at different characteristic scales. A linear model can be set up to split the stochastic process representing the phenomenon into several uncorrelated stochastic processes, each with a different variogram or covariance function and characterizing different aspect of the overall phenomenon under investigation.
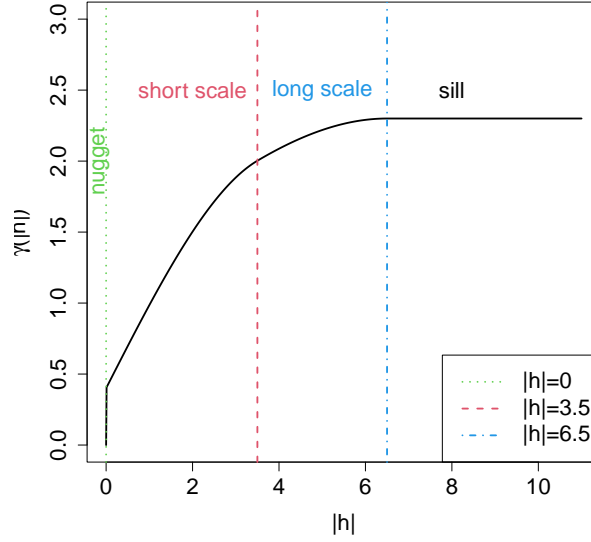
Created on 2024/11/04 at 11:00:38 by Georgios Karagiannis

FIGURE 14.1. Variogram $\gamma(\cdot)$ of $Z(s) = Z_1(s) + Z_2(s) + Z_3(s)$ with spherical s.v. $\gamma_1\left(|h|\,;\sigma^2 = 0.8, \phi = 3.5\right)$, spherical s.v. $\gamma_1\left(|h|\,;\sigma^2 = 1.1, \phi = 6.5\right)$, and nugget $\gamma_3\left(|h|\,;\sigma^2 = 0.4\right)$.

14.1.1. *Decomposition of the random field.*

Note 102. The linear model of regionalization consider the decomposition of the random field of interest $Z(s)$ as a summation of m independent zero-mean random fields $\{Z_j(s)\,;s \in \mathcal{S}\}_{j=0}^m$ each of them characterizing different spatial scales, as

$$(14.1) \qquad\qquad Z(s) = \mu(s) + Z_1(s) + ... + Z_m(s)$$

with $\mu(s) = \mathrm{E}(Z(s))$ be a deterministic drift (or trend) function.

Remark 103. In (14.1), let $Z_j(\cdot)$ be intrinsic random field with semivariogram $\gamma_j(\cdot)$ and mutually independent, then the semivariogram of $Z(\cdot)$ is $\gamma(\cdot) = \sum_{j=1}^m \gamma_j(\cdot)$.

Example 104. For instance consider (14.1) with $\mu(s) = 0$, $m = 3$, $Z_1(s)$ with a spherical semivariogram (11.3) with range $\phi_1 = 3.5$, $Z_2(s)$ with a spherical semi-variogram (11.3) with range $\phi_2 = 6.5$, and $Z_3(s)$ with a nugget semi-variogram. See the "sudden" changes of the line in Figure 14.1 representing change of spatial behavior.

14.1.2. *Scale of variation.*

Note 105. Cressi [1] considers the following intuitive decomposition

$$(14.2) \qquad\qquad Z(s) = \mu(s) + W(s) + \eta(s) + \varepsilon(s), \;\; s \in S$$

where

Created on 2024/11/04 at 11:00:38 by Georgios Karagiannis

$\mu(s) = \mathbf{E}(Z(s))$: is the deterministic mean (or drift) structure. It aims to represent the "large scale variation".

$W(s)$: is a zero mean second order continuous intrinsic random field whose range is larger than gaps between the sites (sampling grid). It aims to represent "smooth small scale variation".

$\eta(s)$: is a zero mean intrinsic random field whose variogram range exists and is smaller than the gaps between the sites. It aims to represent "microscale variation"

$\varepsilon(s)$: is a zero-mean white-noise process (modeled as nugget effect). It aims to represent "measurement error or noise"

$W(s)$, $\eta(s)$, $\varepsilon(s)$ are mutually independent.

Note 106. Reasonably, larger scale components, such as $\mu(s)$, $W(s)$ can represented in the variogram if the diameter of the sampling domain is large $\mathcal{S}$ is large enough.

Note 107. Clearly, smaller scale components, such as $\eta(s)$, $\varepsilon(s)$ could be identified if the sampling grid is sufficiently fine.

Note 108. Decomposition (14.2) is not unique and the components are not clearly identifiable from the data when modeled; e.g. one may find two pairs of $\mu(s), W(s)$ doing the same thing; yet, separating $\eta(s)$ and $\varepsilon(s)$ is difficult as they often describe changes with range smaller than that of the sites (!)

Note 109. The geostatistical model is often presented (with reference to (14.2)) is a form

$$Z(s) = \mu(s) + w(s) + \varepsilon(s), \ s \in S$$

where $w(s) = W(s) + \eta(s)$ contains all the spatial variation.

Note 110. Alternatively, the hierarchical statistical model (Handout 1, 3.5) is used

(14.3) $$Z(s) = Y(s) + \varepsilon(s), \ s \in S$$

where $Y(s) = \mu(s) + W(s) + \eta(s)$ is the spatial model, signal random field or noiseless random field.

Note 111. Another decomposition we will use

$$Z(s) = \mu(s) + \delta(s), \ s \in S$$

where $\delta(s) = W(s) + \eta(s) + \varepsilon(s)$ is the called the correlated process.

Note 112. In several problems, additional covariates may be considered. The available dataset is of the form $\{(x_i, s_i, Z_i)\}_{i \in \mathcal{S}}$ where $Z_i := Z(s_i, x_i)$ is the observed response at

location s_i, associated with the p-dimensional covariate $x_i = (x_{i,1}, ..., x_{i,p})^\top$ for $i \in \mathcal{S}$. Although not a necessity rule, the effect of the associated p-dimensional covariates is often expressed via the deterministic drift function $\mu(s, x) = \mathrm{E}(Z(s, x))$. E.g. in decomposition in (14.2)

$$(14.4) \qquad Z(s, x) = \mu(s, x) + W(s) + \eta(s) + \varepsilon(s), \ s \in S, x \in \mathcal{X}.$$

Here, to simplify the presentation, we suppress dependence on possible covariates $x \in \mathcal{X}$.

15. Learning the semivariogram

Note 113. Consider a random field $(Z(s) ; s \in \mathcal{S})$, $\mathcal{S} \in \mathbb{R}^d$ observed at n sites $S = \{s_1, ..., s_n\}$, and hence a dataset $\{(s_i, Z(s_i))\}_{i=1}^n$.

Note 114. Consider a decomposition

$$Z(s) = \mu(s) + \delta(s), \ s \in S$$

where $\mu(\cdot)$ is an unknown deterministic drift and $\delta(\cdot)$ is a zero mean intrinsic random field.

Example 115. (Meuse river data set) The Meuse river dataset set, used as a running example gives locations and topsoil heavy metal concentrations, along with a number of soil and landscape variables at the observation locations, collected in a flood plain of the river Meuse, near the village of Stein (NL). Here, we use the topsoil zinc concentration, mg kg-1 soil ("ppm") as quantity of interest (Z). See Figure 15.1a. This is the R dataset `meuse{sp}`.

Example 116. (Wolfcamp-aquifer dataset) We also consider the Wolfcamp-aquifer dataset in Exercise 2 in the Exercise sheet. See Figure 15.2a

15.1. **The semvariogram cloud.**

Assumption 117. *Assume that $(Z(s) ; s \in \mathcal{S})$ in an intrinsic random field with unknown constant mean; aka $Z(s) = \mu + \delta(s)$.*

Definition 118. Dissimilarity between pairs of data values $Z(s_a)$ and $Z(s_b)$ is called the measure

$$(15.1) \qquad \gamma^\star(s_a, s_b) = \frac{1}{2}(Z(s_b) - Z(s_a))^2$$

Definition 119. If we let dissimilarity between pairs of data values $Z(s)$ and $Z(s_b)$ depend on the separation $h = s_b - s$ (lag or orientation) then we get

$$\gamma^\star(h) = \frac{1}{2}(Z(s + h) - Z(s))^2.$$

(A) Bubble plot of zinc (origial scale) vs location

(B) Variogram cloud

(C) Empirical estimate (isotropic)

(D) Empirical estimate (90°)

(E) Empirical estimate (45°)

(F) Empirical estimate (different directions)
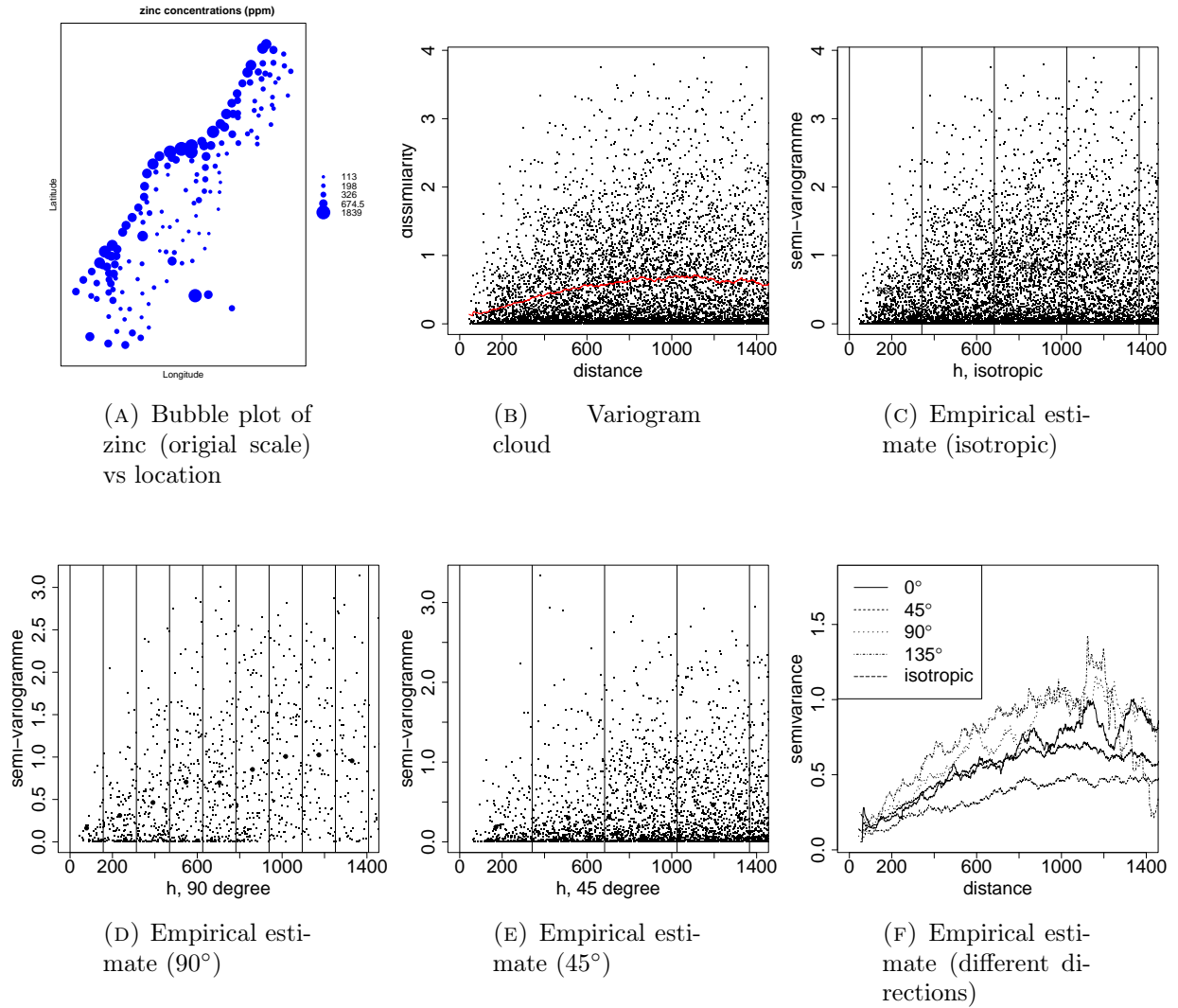
FIGURE 15.1. Meuse dataset variogram estimations (Zinc in log scale)

Definition 120. The semivariogram cloud is the set of $n(n-1)/2$ points

$$\mathfrak{C}_S = \left\{ \left(\left\| s_i - s_j \right\|, \gamma^\star \left(s_j, s_i \right) \right), \ i, j = 1, ..., n, \text{ and } s_i \neq s_j \right\}$$

Note 121. Note that (15.1) is an unbiased estimator of the semivariogram and hence the semivariogram cloud is too.

Note 122. Often there is a smoothing of the cloud is superimposed onto the cloud itself to help us see semivariogram's characteristics (e.g., sill, nugget, range) which may be "hidden" due to potential outliers in the plot.

Created on 2024/11/04 at 11:00:38

by Georgios Karagiannis

(A) Data

(B)   Variogram cloud

(C) Empirical estimate (isotropic)

(D) Empirical estimate (90°)

(E) Empirical estimate (45°)

(F) Empirical estimate (different directions)
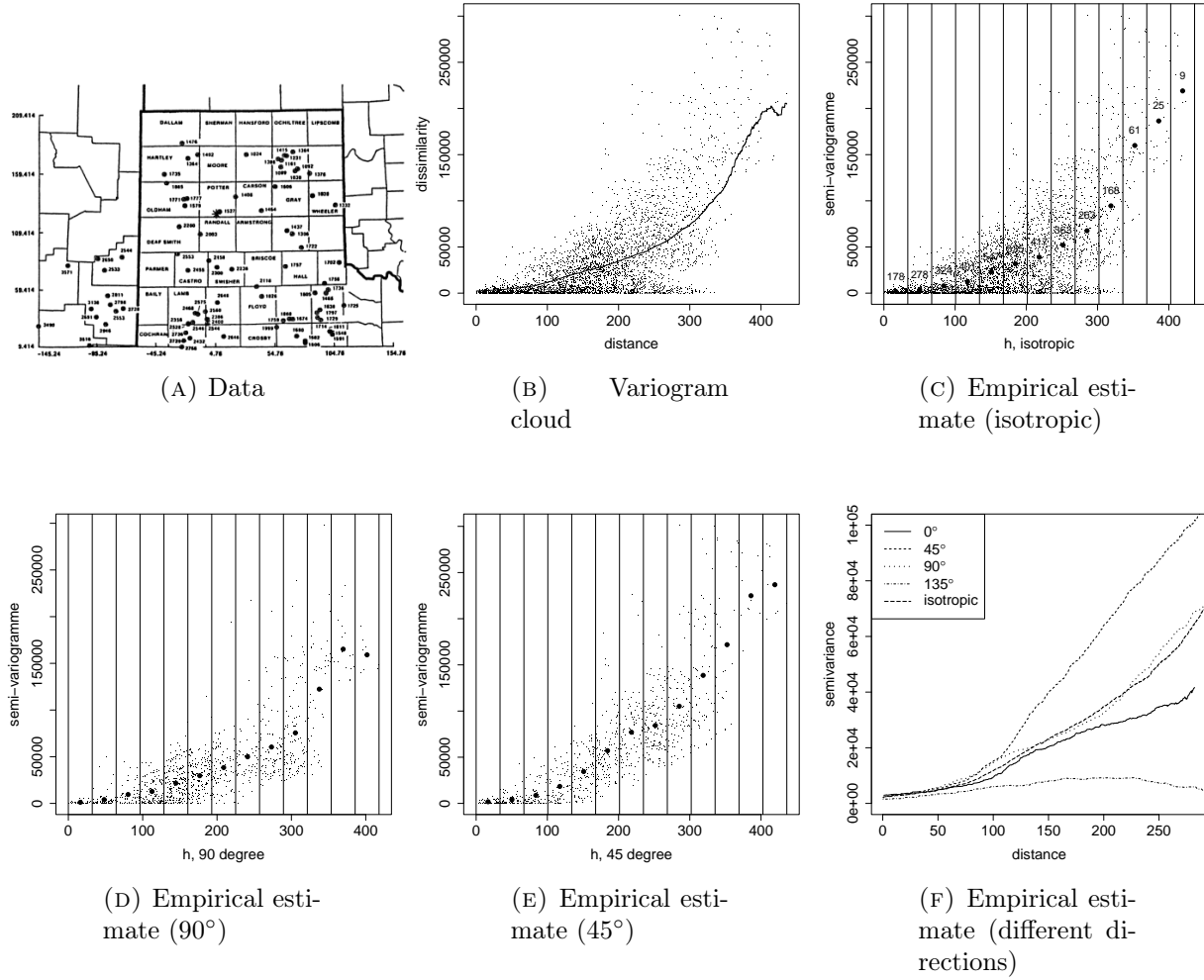
FIGURE 15.2. Wolfcamp-aquifer dataset variogram estimations

Example 123. Figure 15.1b and Figure 15.2b show the semivariogram cloud plots (that is a point plot of the dissimilarities vs the distances) for the datasets Meuse and Wolfcamp-aquifer dataset. The red line is a smoother line of the cloud.

15.2. **Non-parametric semivariogram estimator of $\gamma(\cdot)$.**

Assumption 124. *Assume that $(Z(s); s \in \mathcal{S})$ in an intrinsic random field with unknown constant mean; aka $\mu(\cdot)$ is an unknown constant.*

Proposition 125. *The Smoothed Matheron estimator $\hat{\gamma}(\cdot)$ of semivariogram $\gamma(\cdot)$ of an unknown constant mean intrinsic random field $Z(\cdot)$ is*

$$(15.2) \qquad \hat{\gamma}_M(h) = \frac{1}{2\left|N_{r_1, r_2}(h)\right|} \sum_{\forall (s_i, s_j) \in N_{r_1, r_2}(h)} (Z(s_i) - Z(s_j))^2$$

Created on 2024/11/04 at 11:00:38   by Georgios Karagiannis

where

$$N_{r_1,r_2}(h) = \{(s_i, s_j) \in \mathcal{S} : s_i - s_j \in B_{r_1,r_2}(h)\}$$

contains all the pairs of spatial points points whole difference is in a ball

(15.3) $$B_{r_1,r_2}(h) = \left\{ x : |\|x\| - \|h\|| < r_1, \text{ and } \left\| \frac{x}{\|x\|_2} - \frac{h}{\|h\|_2} \right\|_2 < r_2 \right\}$$

centered at h with radius $r_1 > 0$ and $r_2 > 0$.

Note 126. If we consider isotropic semivariogram $\gamma(\cdot)$ then the ball may just considerate only the length of the distance as

(15.4) $$B_{r_1}(h) = \{ x : |\|x\| - \|h\|| < r_1 \}$$

because the direction does not have any effect.

Note 127. The choice of r_1, r_2 is an art, and a trade-off between variance and bias, similar to the bin length in histograms.

Note 128. In practice, we consider a finite number of k separations $\mathcal{H} = \{h_1, ..., h_k\}$, we estimate in such a way that each class contains at least 30 pairs of points. Then compute $\{\hat{\gamma}_{\mathrm{M}}(h); h \in \mathcal{H}\}$, and plot $\{(h_j, \hat{\gamma}_{\mathrm{M}}(h_j)); j = 1, ..., k\}$.

Example 129. Figures 15.1c and 15.2c, show the nonparametric estimator ignoring the direction for the datasets Meuse and Wolfcamp-aquifer dataset. The estimator is calculated by using the ball in (15.4).

Example 130. Figures 15.2dand 15.1e show the nonparametric estimator considering directions 90° and 45° for the dataset Meuse. Figures 15.2d and 15.2e do the same for the Wolfcamp-aquifer dataset. The estimator is calculated by using the ball (15.3).

Note 131. In practice anisotropies are detected by inspecting experimental semivariograms in different directions and are induced into the model by tuning predefined anisotropy parameters.

Example 132. Figure 15.1f and 15.2a show the nonparametric semivariogram estimator for different directions for the two datasets. We observe possible anisotropy due to the differences in the lines.

15.3. **Classic parametric estimator of $\gamma(\cdot)$.**

Assumption 133. *Consider (for now) the assumption that $(Z(s); s \in \mathcal{S})$ in an intrinsic random field with unknown constant mean; aka $\mu(\cdot)$ is an unknown constant.*

Note 134. Smoothed Matheron estimator (15.2) 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 135. Popular parametrized isotopic semivariogrames are those Section 11.1. Anisotropic semi-variogrames/covariogrames can be specified by using isotropic ones and applying a rotation and dilation as in Example 86.

Note 136. Below are some properties that allow the specification of sophisticated semivariograms from simpler ones.

(1) $\tilde{\gamma}(h) = \gamma(Ah)$ where $\gamma(\cdot)$ is a semivariogram and A constant matrix.
(2) $\gamma(\cdot) = \sum_{i=1}^{n} a_i \gamma_i(\cdot)$, if $a_i \geq 0$, and $\{\gamma_i(\cdot)\}$ are semivariograms
(3) $\gamma(\cdot) = \prod_{i=1}^{n} \gamma_i(\cdot)$, if $\{\gamma_i(\cdot)\}$ are semivariograms
(4) $\gamma(\cdot) = \int \gamma_u(\cdot)\,dF(u)$, if $\gamma_u(\cdot)$ is a semivariogram parametrized by $u \sim F$
(5) $\gamma(\cdot) = \lim_{n\to\infty} \gamma_n(\cdot)$ if $\gamma_n(\cdot)$ is semivariogram and the limit exists
(6) $\gamma_Z(h) = \gamma_Y(h) + \gamma_X(h)$ corresponds to random field $Z(s) = Y(s) + X(s)$ if $(Y(s) : s \in \mathcal{S})$ and $(X(s) : s \in \mathcal{S})$ are independent intrinsic random fields with semivariograms $\gamma_Y(\cdot)$ and $\gamma_X(\cdot)$.
(7) $\gamma(\cdot)$ is a semivariogram iff $\exp(-a\gamma(\cdot))$ is positive definite for any $a > 0$.

Example 137. $\gamma(h) = \|h\|^2$ is a semivariogram because $\exp\left(-a\|h\|_2^2\right)$ is a c.f. for any $a > 0$ and hence positive definite.

Note 138. For a q.m. continuous $(Z(s) : s \in \mathcal{S})$, it is $\lim_{\|h\|\to 0} \gamma(h) = 0$ 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 $\lim_{\|h\|\to 0} \gamma(h) = \sigma_\varepsilon^2 \neq 0$ aka a nugget. Nugget effect is often mathematically described by considering a decomposition ; [5; Ch 1.4.1]

$$Z(s) = Y(s) + \varepsilon(s)$$

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

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

Note 139. Let $\hat{\gamma}$ be the empirical semivariogram $\hat{\gamma}$ (e.g., Matheron (15.2)) computed at k classes, i.e. it is available $\{h_j, \hat{\gamma}(h_j)\}_{j=1}^{k}$. Let γ_θ be a parametrised semivariogram by the unknown θ. The Least Square Errors (LSE) estimator is $\hat{\gamma}_{\text{LSE}}(h) = \gamma\left(h; \hat{\theta}_{\text{LSE}}\right)$ where

(15.5)
$$\hat{\theta}_{\text{LSE}} = \arg\min_\theta (\hat{\gamma} - \gamma(h;\theta))^\top V(\theta)(\hat{\gamma} - \gamma(h;\theta)),$$

Created on 2024/11/04 at 11:00:38 by Georgios Karagiannis

$V(\theta)$ is a user specific positive definite matrix $V(\theta)$ serving as a weight, $\hat{\gamma} = (\hat{\gamma}(h_1), ..., \hat{\gamma}(h_k))^\top$, and $\gamma(h; \theta) = (\gamma(h_1; \theta), ..., \gamma(h_k; \theta))^\top$.

Note 140. An example is OLS

$$\hat{\theta}_{\text{OLS}} = \arg\min_\theta \left(\sum_j (\hat{\gamma}(h_j) - \gamma(h; \theta))^2 \right) \tag{15.6}$$

Example 141. Figures 15.3a and 15.3b show the OLE and WLE estimates (15.6) and (18.1) of the exponential and spherical semivariogram for the Meuse dataset. Figure 15.3c shows the OLE and WLE estimates (15.6) and (18.1) of the exponential semi-variogram for the Wolfcamp dataset. The parametric semivariograms were tuned against the non-parametric estimator (15.2) presented in dots, as discussed in Proposition 139.

15.4. **Parametric learning of nonzero $\mu(\cdot)$ and $\gamma(\cdot)$.**

Note 142. Assume a random field model $(Z(s); s \in \mathcal{S})$ decomposed as

$$Z(s) = \mu(s) + \delta(s)$$

where the trend $\mu(s)$ is parameterized as $\mu(s) = \mu(s; \beta)$ with unknown β (e.g. $\mu(s; \beta) = \psi^\top(s)\beta$), and the zero mean intrinsic process $\delta(s)$ has a semivariogram $\gamma(h)$ parameterised as $\gamma(h) = \gamma(h; \theta)$ with unknown θ.

15.4.1. *Non-parametric inference.*

Note 143. Semi-parametric learning is as follows:

(1) Compute estimate $\hat{\beta}$ via LSE (or equivalent)

$$\hat{\beta}_{\text{LSE}} = \arg\min_\theta \left(\sum_i (Z(s_i) - \mu(s_i; \beta))^2 \right) \tag{15.7}$$

(2) Compute the residuals $\hat{\delta} := \hat{\delta}(s_i)$ from

$$\hat{\delta}(s_i) = Z(s_i) - \mu\left(s_i; \hat{\beta}_{\text{LSE}}\right) \tag{15.8}$$

(3) Compute empirical variogram against $\hat{\delta}$ on $\mathscr{H}$ according to Proposition 125.
(4) Compute estimates $\hat{\theta}_{\text{LSE}}$ and $\hat{\gamma}_{\text{LSE}}(h)$ according to Proposition 139.

Example 144. Figure 15.3a and 15.3b fit an exponential c.f. and a spherical c.f. in the data of Meuse dataset (assuming constant mean); we cannot eyeball any big difference. Figure 15.3c fit an exponential c.f. in the data of Wolfcamp dataset (assuming constant mean); the fit looks really bad, possibly we should consider a non-constant mean and remove the trend.

(A) Exponential c.f. with constant mean (Meuse dataset)

(B) Spherical c.f. with constant mean (Meuse dataset)

(C) Exponential c.f. with constant mean (Wolfcamp dataset)

(D) Exponential c.f. with linear trend (Meuse dataset)

(E) Exponential c.f. with linear trend (Wolfcamp dataset)
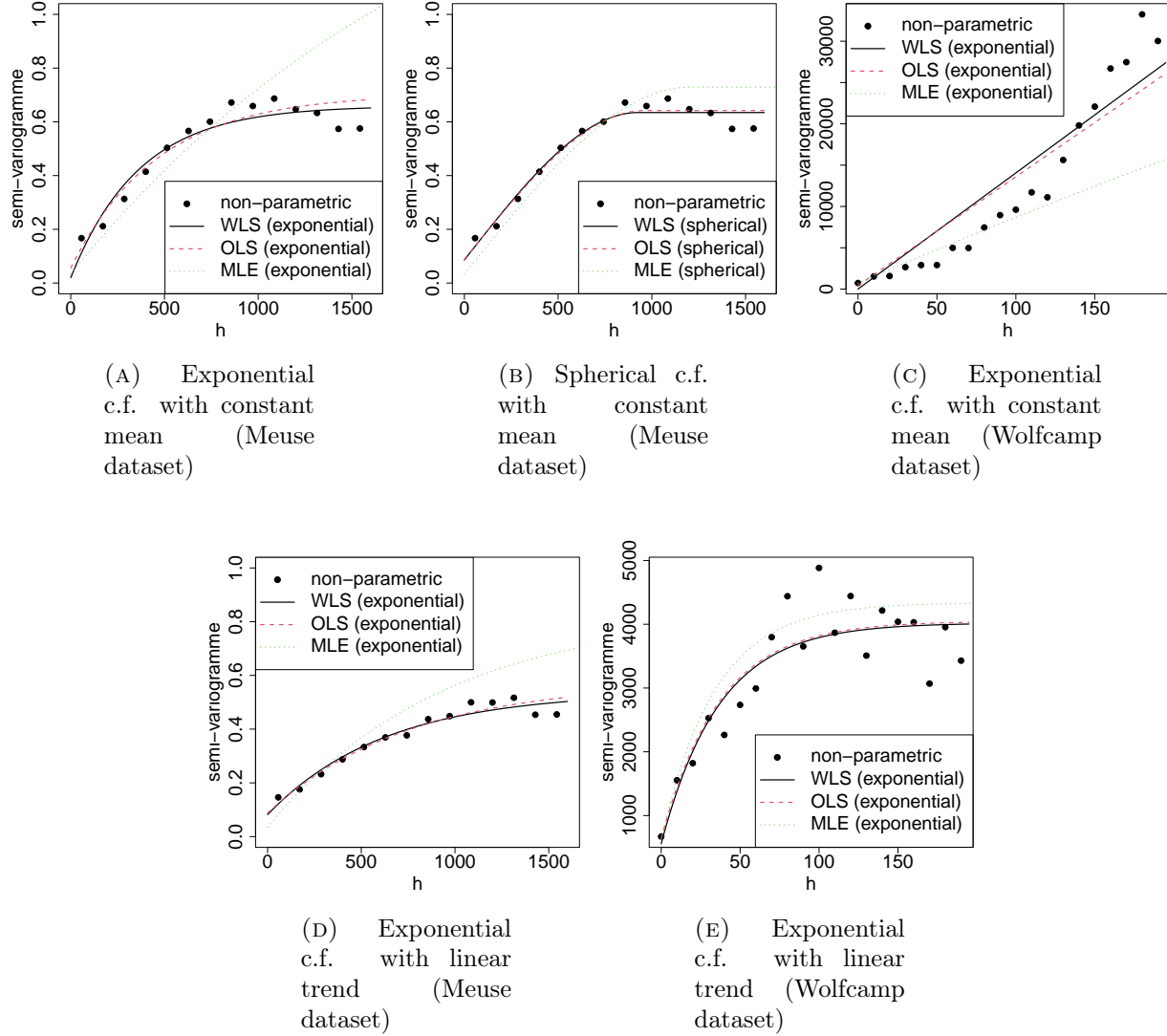
FIGURE 15.3. Parametric training

Example 145. Figure 15.3d fits an exponential c.f. in the residuals $\delta(s) = Z(s) - \mu(s)$ where $\mu(s) = \beta_0 + \beta_1 s_1 + \beta_2 s_2$ and $\hat{\beta}_{\mathrm{OLS}} = (-42.8, -9.5 \cdot 10^{-4}, -6.6 \cdot 10^{-4})^{\top}$ in Meuse dataset. Possibly inference would suggest a constant mean function. Figure 15.3e fits an exponential c.f. in the residuals $\delta(s) = Z(s) - \mu(s)$ where $\mu(s) = \mu(s) = \beta_0 + \beta_1 s_1 + \beta_2 s_2$ and $\hat{\beta}_{\mathrm{OLS}} = (-607, -1.12, -1.13)^{\top}$ in Wolfcamp dataset; we see an improvement in fit compared to Figure 15.3c.

15.4.2. *Parametric inference via MLE.*

Note 146. Assume that the probability distribution of the random field $(Z(s); s \in \mathcal{S})$ is known. The MLE estimates $\left(\hat{\beta}_{\text{MLE}}, \hat{\theta}_{\text{MLE}}\right)$ of (β, θ) can be computed as

$$\left(\hat{\beta}_{\text{MLE}}, \hat{\theta}_{\text{MLE}}\right) = \arg\min_{(\beta, \theta)} \left(-2\log\left(L\left(z_1, ..., z_n | \beta, \theta\right)\right)\right)$$

where $L\left(z_1, ..., z_n | \beta, \theta\right)$ is the associated likelihood function given observed data $\{(s_i, Z_i)\}_{i=1}^n$.

Example 147. If s $Z(\cdot) \sim \text{GP}\left(\mu\left(\cdot; \beta\right), c\left(\cdot, \cdot; \sigma^2, \phi\right)\right)$, with $\mu\left(s; \beta\right) = \beta_0 + s_1\beta_1 + s_2\beta_2$ and $\gamma_{(\sigma^2, \phi)}\left(h\right) = \sigma^2\left(1 - \exp\left(-\phi h^2\right)\right)$ then MLE of (β, σ^2, ϕ) is

$$\left(\hat{\beta}_{\text{MLE}}, \hat{\sigma}^2_{\text{MLE}}, \hat{\phi}_{\text{MLE}}\right) = \arg\min_{\beta, \sigma^2, \phi} \left(-2\log\left(\text{N}\left(Z | \mu_\beta, C_{\sigma^2, \phi}\right)\right)\right)$$

where $\text{N}\left(Z | \mu_\beta, C_\theta\right)$ is the Gaussian pdf at $Z = \left(Z\left(s_1\right), ..., Z\left(s_n\right)\right)^\top$, with mean $[\mu_\beta]_i = \mu\left(s_i; \beta\right) = \beta_0 + s_{1,i}\beta_1 + s_{2,i}\beta_2$ and covariance matrix $[C_{\sigma^2, \phi}]_{i,j} = \sigma^2 \exp\left(-\phi\left(s_i - s_j\right)^2\right)$.

16. (CLASSICAL) KRIGING (FOR PREDICTION)

Note 148. "Kriging" is a general technique for deriving an estimator / predictor of $Z(\cdot)$ (or a function of it) at a location (such as a spatial point s_0, or a block of points $\{s_j^*\}$ or a subregion v_0) of a spatial region $\mathcal{S}$ by properly averaging out data in the neighborhood around the location of interest.

16.1. **Universal Kriging.**

Note 149. Consider the statistical model specified as a stochastic process $(Z(s); s \in \mathcal{S})$ with

(16.1) $$Z(s) = \mu(s) + \delta(s)$$

where $\mu(s)$ is a deterministic linear expansion of known basis functions $\{\psi_j(\cdot)\}_{j=0}^p$ and unknown coefficients $\{\beta_j\}_{j=0}^p$ such as

$$\mu(s) = \sum_{j=0}^p \psi_j(s)\beta_j = (\psi(s))^\top \beta$$

with $\beta = (\beta_0, ..., \beta_p)^\top$ and $\psi(s) = (\psi_0(s), ..., \psi_p(s))^\top$. Also , $\delta(s)$ is a zero mean random field.

Note 150. Consider an available a dataset $\{(s_i, Z_i)\}_{i=1}^n$ with $Z_i := Z(s_i)$ being a realization of $(Z(s); s \in \mathcal{S})$ at site s_i. At these dataset points (16.1) is vectorized as

$$Z = \mu + \delta = \Psi\beta + \delta$$

with vectors $Z = \left(Z\left(s_1\right), ..., Z\left(s_n\right)\right)^\top$, $\delta = \left(\delta\left(s_1\right), ..., \delta\left(s_n\right)\right)^\top$, and $\mu = \left(\mu\left(s_1\right), ..., \mu\left(s_n\right)\right)^\top$, and with (design) matrix Ψ such as $[\Psi]_{i,j} = \psi_j(s_i)$.
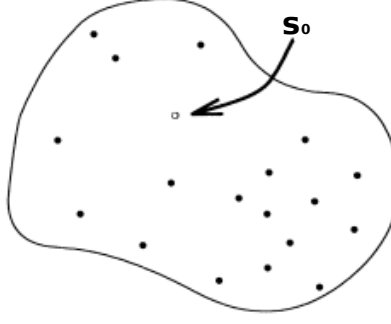
FIGURE 16.1. Kriging area

Note 151. Interested lies in learning/predicting $Z(s_0)$ at an unseen spatial location s_0 within the spatial domain $\mathcal{S}$ (Figure 16.1).

Note 152. "Universal Kriging" (UK) is the technique for producing a Best Linear Unbiased Estimator (BLUE) predictor for $Z_0 := Z(s_0)$ at spatial location $s_0 \in \mathcal{S}$ as a weighted average of the available data around in a neighborhood around that location .

Definition 153. The Universal Kriging (UK) predictor $Z_{\mathrm{UK}}(s_0)$ of $Z(s_0)$ at location $s_0 \in \mathcal{S}$ is the Best Linear Unbiased Estimator (BLUE) of $Z(s_0)$ given the data $\{(s_i, Z_i)\}_{i=1}^n$.

Note 154. [**L**INEAR] The UK predictor $Z_{\mathrm{UK}}(s_0)$ of $Z(s_0)$ at s_0 has the following linear form weighted by a set of tunable unknown weights $\{w_i\}$

$$(16.2) \qquad Z_{\mathrm{UK}}(s_0) = w_{n+1} + \sum_{i=1}^n w_i Z(s_i)$$

$$(16.3) \qquad = w_{n+1} + w^\top Z$$

where $Z = (Z_1, ..., Z_n)^\top$ and $w = (w_1, ..., w_n)^\top$.

Note 155. [**U**nbiased] For (16.2), to satisfy unbiasness (that is zero systematic error"), we need

$$\mathrm{E}(Z_{\mathrm{UK}}(s_0)) = \mathrm{E}(Z(s_0)) \Leftrightarrow w_{n+1} + \sum_{i=1}^n w_i \mathrm{E}(Z(s_i)) = \mu(s_0)$$

$$\Leftrightarrow w_{n+1} + \sum_{i=1}^n w_i \mu(s_i) = \mu(s_0) \Leftrightarrow w_{n+1} + \sum_{i=1}^n w_i (\psi(s_i))^\top \beta = (\psi(s_0))^\top \beta$$

$$(16.4) \qquad \Leftrightarrow w_{n+1} + w^\top \Psi \beta = \Psi_0 \beta$$

where Ψ is matrix with $[\Psi]_{i,j} = \psi_j(s_i)$ and Ψ_0 is a (column) vector with $[\Psi_0]_{1,j} = \psi_j(s_0)$. Because in (16.4) both sides are polynomial w.r.t β all coefficients must be equal; hence

Created on 2024/11/04 at 11:00:38 by Georgios Karagiannis

sufficient and necessary conditions for unbiaseness are

$$(16.5) \qquad \text{Assumption:} \qquad (\psi(s_0))^\top = \sum_{i=1}^n w_i (\psi(s_i))^\top \qquad \Leftrightarrow \Psi_0 = w^\top \Psi$$

$$(16.6) \qquad \text{Assumption:} \qquad w_{n+1} = 0$$

Note 156. We set $\psi_0(\cdot) = 1$; then (16.4) implies

$$(16.7) \qquad \text{Assumption:} \qquad \sum_{i=1}^n w_i = 1 \qquad \Leftrightarrow w^\top \underline{1} = 1$$

Note 157. The MSE of $Z_{\text{UK}}(s_0)$, given some of the above Assumptions, is

$$(16.8) \qquad \text{MSE}(Z_{\text{UK}}(s_0)) = \text{E}(Z_{\text{UK}}(s_0) - Z(s_0))^2$$

$$= \text{E}(\Psi\beta + \delta(s_0) - w^\top \Psi\beta - w^\top \delta)^2; \qquad \left\{ \text{let } \delta = (\delta(s_1), ..., \delta(s_n))^\top \right\}$$

$$(16.9) \qquad = \text{E}\left(\sum_{i=1}^n w_i \delta(s_i) - \delta(s_0) \right)^2 \overset{\sum_{i=1}^n w_i = 1}{=} \text{E}\left(\sum_{i=1}^n w_i (\delta(s_i) - \delta(s_0)) \right)^2$$

$$(16.10) \qquad = -\text{E}\left(\frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n w_i w_j (\delta(s_i) - \delta(s_j))^2 - 2\frac{1}{2} \sum_{i=1}^n w_i (\delta(s_i) - \delta(s_0))^2 \right)$$

$$(16.11) \qquad = -\sum_{i=1}^n w_i \sum_{j=1}^n w_j \frac{1}{2} \text{E}(\delta(s_i) - \delta(s_j))^2 + 2 \sum_{i=1}^n w_i \frac{1}{2} \text{E}(\delta(s_i) - \delta(s_0))^2$$

Note 158. To impose tractability, we consider the following additional assumption

$$(16.12) \quad \text{Assumption:} \qquad (\delta(s); s \in \mathcal{S}) \text{ intrinsic random field } \text{ with semivariogram } \gamma(\cdot)$$

Note 159. Since $(\delta(s); s \in \mathcal{S})$ is intrinsic stationary, and its semivariogram exists, $\text{MSE}(Z_{\text{UK}}(s_0))$ can be expressed w.r.t. the semivariogram as

$$(16.13) \qquad \text{MSE}(Z_{\text{UK}}(s_0)) = -\gamma(s_0 - s_0) - \sum_{i=1}^n w_i \sum_{j=1}^n w_j \gamma(s_i - s_j) + 2 \sum_{i=1}^n w_i \gamma(s_i - s_0)$$

$$(16.14) \qquad = -\gamma_{00} - w^\top \mathbf{\Gamma} w + 2w^\top \boldsymbol{\gamma}_0$$

where $w = (w_1, ..., w_n)^\top$, $\boldsymbol{\gamma}_0 = (\gamma(s_0 - s_i), ..., \gamma(s_0 - s_n))^\top$, $\gamma_{00} = \gamma(s_0 - s_0)$, and $[\mathbf{\Gamma}]_{i,j} = \gamma(s_i - s_j)$.

Note 160. [**B**est] The Lagrange function for minimizing the MSE (16.13) under (16.4) is

$$\mathfrak{L}(w, \lambda) = -\sum_{i=1}^n \sum_{j=1}^n w_i w_j \gamma(s_i - s_j) + 2 \sum_{i=1}^n w_i \gamma(s_0 - s_i) - \sum_{j=0}^p \lambda_j \left(\sum_{i=1}^n w_i \psi_j(s_i) - \psi_j(s_0) \right)$$

$$= -w^\top \mathbf{\Gamma} w + 2w^\top \boldsymbol{\gamma}_0 - (w^\top \Psi - \Psi_0) \lambda$$

Note 161. The UK system of equations is

$$0 = \nabla_{(w,\lambda)} \mathfrak{L}(w,\lambda)\big|_{(w_{\mathrm{UK}},\lambda_{\mathrm{UK}})} \iff$$

(16.15)
$$\begin{cases} 0 = -2\sum_{j=1}^{n} w_{\mathrm{UK},j}\gamma(s_i - s_j) + 2\gamma(s_0 - s_i) - \sum_{j=0}^{p}\lambda_{\mathrm{UK},j}\psi_j(s_i), \\ \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad i = 1,...,n \iff \\ \psi_j(s_0) = \sum_{i=1}^{n} w_{\mathrm{UK},i}\psi_j(s_i), \quad j = 0,...,p \end{cases}$$

(16.16)
$$\begin{cases} 0 = -2\mathbf{\Gamma}w + 2\boldsymbol{\gamma}_0 - \Psi\lambda_{\mathrm{UK}} \\ \Psi_0 = w_{\mathrm{UK}}^\top \Psi \end{cases}$$

Then by multiplying both sides by $\Psi^\top \Gamma^{-1}$ I get

$$0 = -2\Psi^\top\Gamma^{-1}\mathbf{\Gamma}w_{\mathrm{UK}} + 2\Psi^\top\Gamma^{-1}\boldsymbol{\gamma}_0 - \Psi^\top\Gamma^{-1}\Psi\lambda_{\mathrm{UK}} \iff$$

(16.17)
$$\lambda_{\mathrm{UK}} = 2\left(\Psi^\top\Gamma^{-1}\Psi\right)^{-1}\left(\Psi^\top\Gamma^{-1}\boldsymbol{\gamma}_0 - \Psi_0^\top\right)$$

Then by substituting (16.17) in (16.16), I get the UK weights as

(16.18)
$$w_{\mathrm{UK}} = \mathbf{\Gamma}^{-1}\left(\boldsymbol{\gamma}_0 - \Psi\left(\Psi^\top\Gamma^{-1}\Psi\right)^{-1}\left(\Psi^\top\Gamma^{-1}\boldsymbol{\gamma}_0 - \Psi_0^\top\right)\right)$$

Note 162. Then by substituting 16.18 in (19.2) , the UK predictor $Z_{\mathrm{UK}}(s_0)$ at s_0 is

(16.19)
$$Z_{\mathrm{UK}}(s_0) = \left(\boldsymbol{\gamma}_0 - \Psi\left(\Psi^\top\Gamma^{-1}\Psi\right)^{-1}\left(\Psi^\top\Gamma^{-1}\boldsymbol{\gamma}_0 - \Psi_0^\top\right)\right)^\top \mathbf{\Gamma}^{-1}Z$$

and by substituting 16.18 in (16.14) its standard error is

(16.20)
$$\sigma_{\mathrm{UK}}(s_0) = \sqrt{-w_{\mathrm{UK}}^\top\mathbf{\Gamma}w_{\mathrm{UK}} + 2w_{\mathrm{UK}}^\top\boldsymbol{\gamma}_0}$$

(16.21)
$$= \sqrt{\boldsymbol{\gamma}_0\Gamma^{-1}\boldsymbol{\gamma}_0 - \left(\Psi^\top\Gamma^{-1}\boldsymbol{\gamma}_0 - \Psi_0^\top\right)^\top\left(\Psi^\top\Gamma^{-1}\Psi\right)^{-1}\left(\Psi^\top\Gamma^{-1}\boldsymbol{\gamma}_0 - \Psi_0^\top\right)}$$

Note 163. $(1-\alpha)\,100\%$ Prediction interval of UK predictor $Z_{\mathrm{UK}}(s_0)$ at s_0 is

(16.22)
$$\left(Z_{\mathrm{UK}}(s_0) - q_{\alpha/2}\sqrt{\sigma_{\mathrm{UK}}^2(s_0)}\,,\; Z_{\mathrm{UK}}(s_0) + q_{1-\alpha/2}\sqrt{\sigma_{\mathrm{UK}}^2(s_0)}\right)$$

where $q.$ are suitable quantiles of the distribution of $Z(\cdot)$. E.g. if $Z(\cdot) \sim \mathrm{GP}\left(\mu(\cdot), c(\cdot,\cdot)\right)$ then $q_{0.05/2} = -1.96$ and $q_{0.95/2} = 1.96$ at $\alpha = 0.05$.

Note 164. Note that we have not assumed a particular distribution of $Z(\cdot)$ or $\delta(\cdot)$, but only stationarity assumptions.

Note 165. Assumption in (21.3) restricts the available models (16.2) under consideration. Essentially, it addresses the non-determination issue caused by the assumption (16.12) (introducing intrinsic stationarity δ) which can only partially characterize the covariance function of δ. Also it ensures conditional negative definiteness for 16.14.

Note 166. It was not necessary to consider the intrinsic stationarity in Note 158 to derive Universal Kriging predictor equations; formulas (16.19) & (16.20) could have been derived with respect to the covariance function $c(\cdot,\cdot)$ of $(\delta(\cdot))$ instead of its semivariogram $\gamma(\cdot)$. Here, intrinsic stationarity was assumed for practical reasons; it allowed us to express 16.19 and (16.20) as functions of the semivariogram whose estimation has been discussed in Section 15.

Note 167. Practical use of (16.19), (16.20), and (16.22) requires knowledge of the linear drift coefficients $\{\beta_j\}$ and the semivariogram $\gamma(\cdot)$. If $\{\beta_j\}$ and $\gamma(\cdot)$ are unknown, a way to learn them, is as follows. Consider a separate dataset training dataset $\{(s'_i, Z'_i)\}_{i=1}^{n'}$ (although in practice we use the same $\{(s_i, Z_i)\}_{i=1}^{n}$). Model the semivariogram $\gamma(\cdot)$ with a conditional negative semi-definite function $\gamma_\theta(\cdot)$ parameterized by an unknown parameter θ. Then train/estimate $\{\beta_j\}$ and $\gamma_\theta(\cdot)$ against dateset $\{(s'_i, Z'_i)\}_{i=1}^{n'}$ by using the semi-parametric procedure in Note 143 or Note 146, and plug the estimates in (16.19), (16.20), and (16.22) .

Note 168. The dataset $\{(s'_i, Z'_i)\}_{i=1}^{n'}$ used to train $\{\beta_j\}$ and $\gamma(\cdot)$ should be different than the dataset $\{(s_i, Z_i)\}_{i=1}^{n'}$ used to produce the Kriging predictive equations (16.19), (16.20), and (16.22). This is because in theory the same dataset should not be used twice (it leads to overconfidence). However often in practice the same dataset is use in violation of the theory because usually this has a small impact in the results.

Example 169. [2] Consider the example with the Meuse dataset. Fig 16.2b presents the UK prediction $Z_{\text{UK}}(s_0)$ at any point $s_0 \in \mathcal{S}$ under model (16.1) for which the deterministic drift mean has a linear form $\mu(s) = \beta_0 + \beta_1 s_1 + \beta_2 s_2$. Following Notes 167 and 143, we computed the $\hat{\beta}_{\text{LSE}}$ of β by (15.7), we computed the residual process $\left\{\hat{\delta}_i\right\}$ by removing the linear trend by (15.8), we computed the non-parametric estimate of semivariogram $\hat{\gamma}$ (15.2) of δ as in Proposition 125. We considered a (parametric) isotropic exponential semivariogram $\gamma_{(\sigma^2,\phi)}$ of δ where we computed the OLS $\hat{\theta}_{\text{OLS}} = \left(\hat{\sigma}^2_{\text{OLS}}, \hat{\phi}_{\text{OLS}}\right)$ of the hyperparameters (σ^2, ϕ) as in (15.6) (see Figure 15.3d).Then we plugged in the estimated $\gamma_{(\hat{\sigma}^2_{\text{OLS}}, \hat{\phi}_{\text{OLS}})}$ in (16.19) to compute the UK weights w_{UK} for the UK predictor $Z_{\text{UK}}(s_0) = w_{\text{UK}} Z$ for any $s_0 \in \mathcal{S}$.

Example 170. (Cont. Examples 115, 132) Consider the example with the Meuse dataset. The dataset has another measurement (a potential regressor in the deterministic mean $\mu(s)$), the "distance to the Meuse river bed" $\{d_i\}$ at the associated locations $\{s_i\}$, let's denote it by d. Figure 16.2c shows a rather linear relationship between Z and $\sqrt{d}$. Consider deterministic drift mean has a linear form $\mu(s, d) = \beta_0 + \beta_1 \sqrt{d(s)}$ while the rest specification is the same

---

as in Example 169. We follow the same procedure as in Example 169 and we get the UK predictor in Figure 16.2d.

16.2. **Ordinary Kriging.**

Note 171. Ordinary Kriging (OK) addresses spatial prediction in cases that the specified statistical model on $(Z(s); s \in \mathcal{S})$ has the form

$$(16.23) \qquad\qquad Z(s) = \beta_0 + \delta(s)$$

with unknown $\beta_0 \neq 0$ and intrinsically stationary process $(\delta(s); s \in \mathcal{S})$.

Note 172. OK can be derived as a special case of the Universal Kriging by setting $p = 0$ and constant spatial mean $\mu(s) = \beta_0$.

Example 173. [The derivation is in (Exercise 17 Exercise sheet).] For demonstration, we mention some key equations of OK

$$(16.24) \qquad \mathfrak{L}(w, \lambda) = \underbrace{-w^\top \boldsymbol{\Gamma} w + 2w^\top \boldsymbol{\gamma}_0}_{=\text{MSE}(Z_{\text{OK}}(s_0))} - \lambda \underbrace{\left(w^\top \mathbf{1} - 1\right)}_{\text{Assumption } \sum_{i=1}^n w_i = 1}$$

The OK system of equations is $0 = \nabla_{(\{w_i\}, \lambda)} L(w, \lambda)\big|_{(w, \lambda)}$ producing

$$(16.25) \qquad \begin{cases} 0 = -2\boldsymbol{\Gamma} w_{\text{OK}} + 2\boldsymbol{\gamma}_0 - \mathbf{1}\lambda \\ w_{\text{OK}}^\top \mathbf{1} = 1 \end{cases}$$

the weights are

$$(16.26) \qquad\qquad w_{\text{OK}} = \boldsymbol{\Gamma}^{-1}\left(\boldsymbol{\gamma}_0 + \frac{1 - \mathbf{1}^\top \boldsymbol{\Gamma}^{-1}\boldsymbol{\gamma}_0}{\mathbf{1}^\top \boldsymbol{\Gamma}^{-1}\mathbf{1}}\mathbf{1}\right)$$

the Kriging standard error of $Z_{\text{OK}}(s_0)$ at s_0 is

$$(16.27) \qquad\qquad \sigma^2_{\text{OK}}(s_0) = \boldsymbol{\gamma}_0^\top \boldsymbol{\Gamma}^{-1}\boldsymbol{\gamma}_0 - \frac{\left(1 - \mathbf{1}^\top \boldsymbol{\Gamma}^{-1}\boldsymbol{\gamma}_0\right)^2}{\mathbf{1}^\top \boldsymbol{\Gamma}^{-1}\mathbf{1}}.$$

Example 174. (Cont. Examples 115, 132, 170) Consider the example with the Meuse dataset. with $\mu(s) = \beta_0 + \beta_1 s_1 + \beta_2 s_2$. We do not see much difference between OK in Figure 16.2a and UK in Figure 16.2b possibly because the slops in the linear trend (mean) of UK are rather small and insignificant (See Example 145).

16.3. **Simple Kriging.**

Note 175. Simple Kriging (SK) addresses spatial prediction in cases that the specified statistical model on $(Z(s); s \in \mathcal{S})$ has the form

$$(16.28) \qquad\qquad Z(s) = \mu(s) + \delta(s)$$

where the deterministic mean $\mu(s)$ is known, and $(\delta(s); s \in \mathcal{S})$ is a weakly stationary process with covariogram $c(\cdot)$.

Example 176. [The derivation is in (Exercise 15 in the Exercise sheet).] It does not require any assumption in the weights such as (16.5) or (16.25).As a supplementary and for demonstration, we mention the SK predictor at s_0 and standard error:

$$Z_{\mathrm{SK}}(s_0) = \mu(s_0) + C_0^\top C^{-1}[Z - \mu]$$

$$\sigma_{\mathrm{SK}} = \sqrt{c(s_0, s_0) - C_0^\top C^{-1} C_0}$$

with $\mu = (\mu(s_1), ..., \mu(s_n))^\top$, $C_0 = (c(s_0 - s_i), ..., \gamma(s_0 - s_n))^\top$, and $[C]_{i,j} = c(s_i - s_j)$.

Example 177. Consider the example with the Meuse dataset. Fig 16.2a presents the OK prediction $Z_{\mathrm{OK}}(s_0)$ at any point $s_0 \in \mathcal{S}$ under model (16.23) that is the UK case (16.1) for when $\mu(s) = \beta_0$. First we computed the non-parametric semivariogram $\hat{\gamma}$ (15.2) as in Proposition 125; then we considered a (parametric) isotropic exponential semi-variogram $\gamma_{(\sigma^2, \phi)}$ where we computed the OLS $\hat{\theta}_{\mathrm{OLS}} = \left(\hat{\sigma}^2_{\mathrm{OLS}}, \hat{\phi}_{\mathrm{OLS}}\right)$ of the hyperparameters (σ^2, ϕ) as in (15.6) (see Figure 15.3a); and then we plugged in the estimated $\gamma_{(\hat{\sigma}^2_{\mathrm{OLS}}, \hat{\phi}_{\mathrm{OLS}})}$ in (16.26) to compute the OK weights w_{OK} for the OK predictor $Z_{\mathrm{OK}}(s_0) = w_{\mathrm{OK}} Z$ for any $s_0 \in \mathcal{S}$.

17. The Bayesian Kriging paradigm (Hierarchical modeling)

17.1. **General framework.**

Note 178. The Bayesian framework provides an elegant solution for taking into account the uncertainty on variogram or covariance parameters.

Note 179. Consider the geostatistical model for $(Z(s); s \in \mathcal{S})$ with a scale decomposition such as in (14.3)

(17.1)
$$Z(s) = \underbrace{\mu(s) + w(s)}_{=Y(s)} + \varepsilon(s), \quad s \in \mathcal{S}$$

where, $\mu(s) = \mathrm{E}(Z(s))$ is an unknown drift function modelling large scale variations, $(w(s); s \in \mathcal{S})$ is a zero mean random field modelling lower scale variation, and $(\varepsilon(s); s \in \mathcal{S})$ is a nugget random field modeling measurement errors. Also $Y(s) = \mu(s) + w(s)$.

Note 180. Consider a dataset $\{(s_i, Z_i)\}_{i=1}^n$ with $Z_i = Z(s_i)$ being a realization of (17.1) at site $s_i \in \mathcal{S}$.

Note 181. Unlike in the traditional kriging framework, in Bayesian kriging, we have to specify a certain probabilistic model on the spatial random fields. Uncertainty can be decomposed

(A) Ordinary Kriging, constant mean, exponential covariogram

(B) Universial Kriging, mean $\mu(s) = \beta_0 + \beta_1 s_1 + \beta_2 s_2$, exponential covariogram

(C) Scatter plot of Z vs $\log(d)$ where d is a third variable representing the distance to the riven bed.

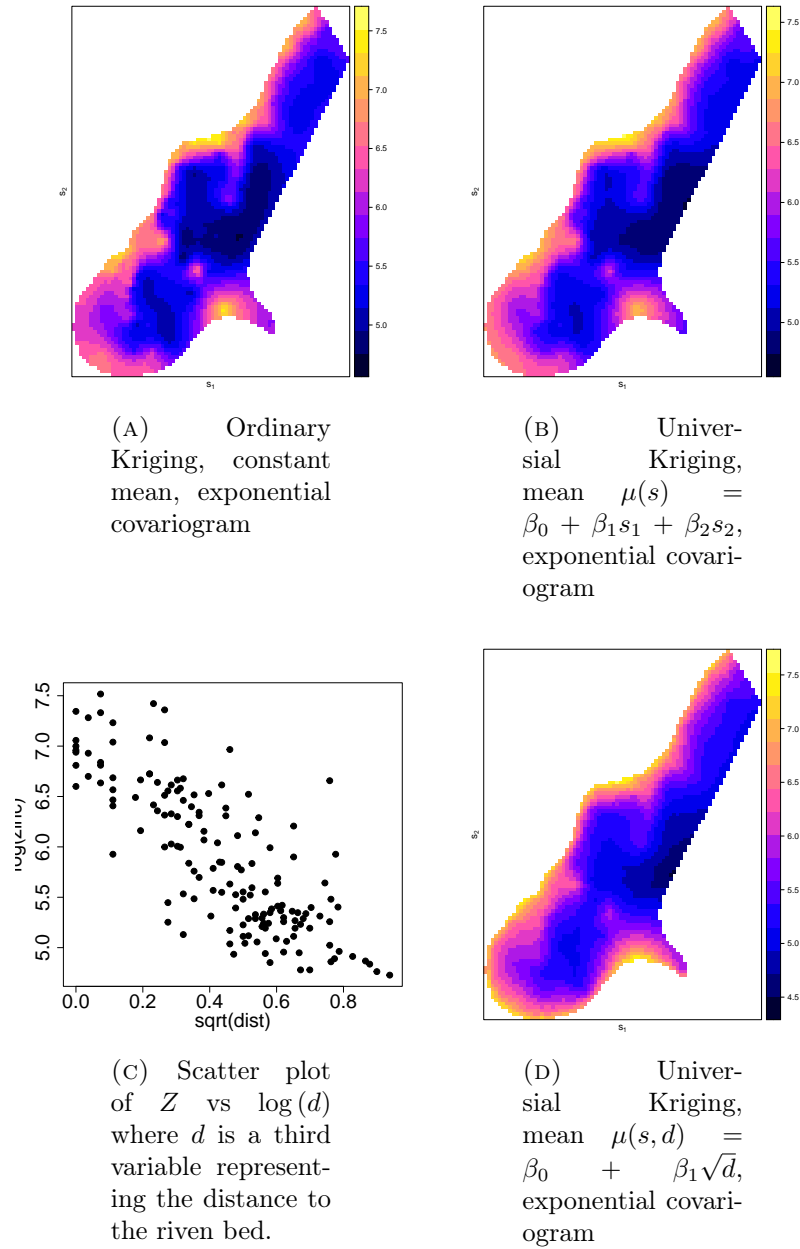(D) Universial Kriging, mean $\mu(s,d) = \beta_0 + \beta_1 \sqrt{d}$, exponential covariogram

FIGURE 16.2. Kriging Meuse dataset.

according to the Hierarchical spatial model

(17.2)
$$\begin{cases} Z|Y,\theta_1,\theta_2 & \text{data model} \\ Y|\theta_1,\theta_2 & \text{spatial process model} \\ \theta_1 & \text{hyper-priormodel (optional layer)} \end{cases}$$
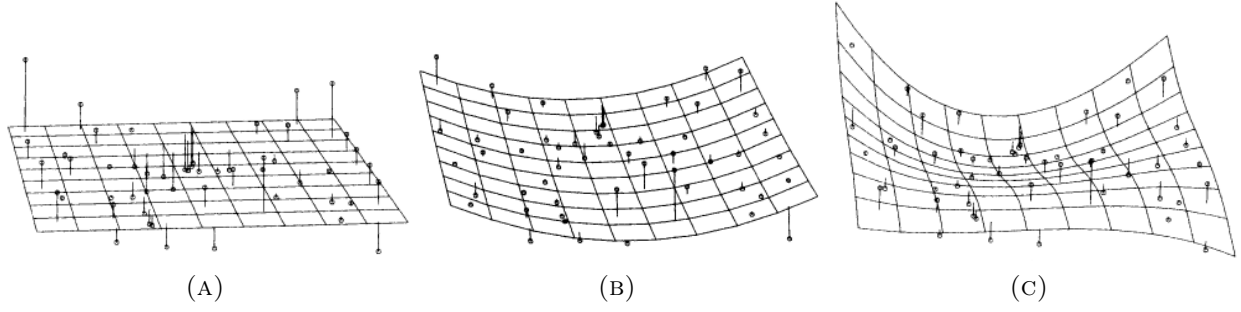
Created on 2024/11/04 at 11:00:38 by Georgios Karagiannis

(A)   (B)   (C)

FIGURE 17.1. Examples representing the hierarchical spatial model 17.2 for different values of ϑ

with

$$\text{pr}\left(Z, Y, \theta_1 | \theta_2\right) = \text{pr}\left(Z | Y, \theta_1, \theta_2\right) \text{pr}\left(Y | \theta_1, \theta_2\right) \text{pr}\left(\theta_1 | \theta_2\right)$$

where $Z = (Z_1, ..., Z_n)^\top$, $Y = (Y_1, ..., Y_n)^\top$, with $Y_i = Y(s_i)$. Here, $\theta_1 \in \Theta_1$ is an unknown random hyper parameter following prior $\theta_1 | \theta_2 \sim \text{pr}\left(\theta_1 | \theta_2\right)$, and $\theta_2 \in \Theta_2$ is an unknown fixed parameter without a specified prior.

Note 182. Spatial process model expresses the scientific uncertainty (e.g., that coming from $(Y(\cdot))$ as quantified via the specified distribution $\text{pr}\left(Y | \theta\right)$ possibly labeled by some hyper-parameter $\theta = (\theta_1, \theta_2)^\top$. Data model expresses the measurement uncertainty (e.g., that coming from $(\varepsilon(\cdot))$) as quantified via the distribution $\text{pr}\left(Z | Y, \theta\right)$ possibly labeled by some parameter θ.

Note 183. Figure 17.1 presents a visualization of the hierarchical model in Notes 181. The surfaces can be considered as a realization of the spatial process model $Y(\cdot)$, and the dotes can be considered as realizations of the data model $\varepsilon(\cdot)$ at specific sites given the spatial process.

Note 184. Under Bayesian model (17.2), unknown but fixed θ_2 can be learned pointwise by computing a point estimator $\hat{\theta}_2$ via marginal likelihood maximization

$$\hat{\theta}_2 = \arg\min_{\theta_2} \left(-2 \log\left(\text{pr}\left(Z | \theta_2\right)\right)\right),$$

$$\text{where} \quad \text{pr}\left(Z | \theta_2\right) = \int \text{pr}\left(Z, Y, \theta_1 | \theta_2\right) \text{d}Y \, \text{d}\theta_1$$

Note 185. Under Bayesian model (7.2), uncertainty about unknown random θ_1 can be represented by the posterior distribution

$$\text{pr}\left(\theta_1 | Z, \theta_2 = \hat{\theta}_2\right) = \frac{\text{pr}\left(Z | \theta_1, \theta_2 = \hat{\theta}_2\right) \text{pr}\left(\theta_1 | \theta_2 = \hat{\theta}_2\right)}{\text{pr}\left(Z | \theta_2 = \hat{\theta}_2\right)}$$

where the value $\hat{\theta}_2$ is plugged in.

Note 186. The posterior predictive distributions of the spatial process model $(Y(\cdot))$ given the data Z is

$$(17.3) \qquad \mathrm{pr}\left(Y(s_0)\,|Z,\theta_2=\hat{\theta}_2\right) = \int \mathrm{pr}\left(Y(s_0)\,|Z,\theta_1,\theta_2=\hat{\theta}_2\right) \mathrm{pr}\left(\theta_1|Z,\theta_2=\hat{\theta}_2\right) \mathrm{d}\theta_1$$

and the marginal process $(Z(\cdot))$ given the data Z is

$$(17.4) \qquad \mathrm{pr}\left(Z(s_0)\,|Z,\theta_2=\hat{\theta}_2\right) = \int \mathrm{pr}\left(Z(s_0)\,|Z,\theta_1,\theta_2=\hat{\theta}_2\right) \mathrm{pr}\left(\theta_1|Z,\theta_2=\hat{\theta}_2\right) \mathrm{d}\theta_1$$

for any $s_0 \in \mathcal{S}$.

Note 187. The Bayes Kriging predictor $\hat{Y}_{\mathrm{BK}}(s_0)$ of $Y(s_0)$ at unseen location s_0 equations is the optimizer

$$(17.5) \qquad \hat{Y}_{\mathrm{BK}}(s_0) = \underset{Y_{\mathrm{BK}}(s_0)}{\arg\min}\left(\mathrm{E}\left(\ell\left(Y_{\mathrm{BK}}(s_0),Y(s_0)\right)|Z\right)\right)$$

given a pre-specified loss function $\ell : \mathcal{Z} \times \mathcal{Z} \to \mathbb{R}_+$. The expectation is under the probability (17.4). For Bayes Kriging predictor $\hat{Z}_{\mathrm{BK}}(s_0)$ of $Z(s_0)$ at s_0 is

$$(17.6) \qquad \hat{Z}_{\mathrm{BK}}(s_0) = \underset{Z_{\mathrm{BK}}(s_0)}{\arg\min}\left(\mathrm{E}\left(\ell\left(Z_{\mathrm{BK}}(s_0),Z(s_0)\right)|Z\right)\right)$$

17.2. **An example: Gaussian process regression.**

Note 188. We are going through a particular example of the Bayesian Kriging (or Bayesian Gaussian process regression) to demonstrate how the "Bayesian Kriging" works.

List of useful formulas.

Fact 189. *Let $X \sim N(\mu_X, \Sigma_X)$, $Y \sim N(\mu_Y, \Sigma_Y)$ and Y, X independent. Let fixed matrices A and B and vector c of appropriate sizes. Then*

$$(17.7) \qquad AX + BY + c \sim N\left(A\mu_X + B\mu_Y + c, A\Sigma_X A^\top + B\Sigma_Y B^\top\right)$$

Fact 190. *Let $N(\beta|b, B)$ be the Gaussian pdf with mean b and covariance B at β. It is*

$$\int N(Z|\Psi\beta, C)\, N(\beta|b, B)\, d\beta = N\left(Z|\Psi b, C + \Psi B \Psi^\top\right)$$

Fact 191. *(Woodbury matrix identity) It is*

$$(A + BCD)^{-1} = A^{-1} - A^{-1}B\left(C^{-1} + DA^{-1}B\right)^{-1}DA^{-1}$$

Fact 192. *[Marginalization & conditioning] Let $x_1 \in \mathbb{R}^{d_1}$, and $x_2 \in \mathbb{R}^{d_2}$. If*

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \sim N_{d_1 + d_2}\left(\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \Sigma_1 & \Sigma_{21}^\top \\ \Sigma_{21} & \Sigma_2 \end{bmatrix}\right)$$

then it is

$$x_2|x_1 \sim N_{d_2}\left(\mu_{2|1}, \Sigma_{2|1}\right)$$

where

$$\mu_{2|1} = \mu_2 + \Sigma_{21}\Sigma_1^{-1}\left(x_1 - \mu_1\right) \text{ and } \Sigma_{2|1} = \Sigma_2 - \Sigma_{21}\Sigma_1^{-1}\Sigma_{21}^\top$$

Note 193. Consider an available dataset $\{(s_i, Z_i)\}_{i=1}^n$ with $\{Z_i \in \mathbb{R}\}$ and $\{s_i \in \mathcal{S}\}$. The i-th datum $Z_i = Z(s_i)$ may be an independent measurement of an unobserved quantity $Y_i = Y(s_i)$ at location s_i but contaminated by additive random error $\varepsilon_i = \varepsilon(s_i)$; i.e. $Z_i = Y_i + \epsilon_i$ for $i = 1, ..., n$. Interest lies in recovering functions $Z(\cdot)$ and/or $Y(\cdot)$ over the spatial domain $\mathcal{S}$.

Specifying the hierarchical model.

Note 194. The geostatistical model can be considered as

$$(17.8) \qquad\qquad Z(s) = \underbrace{\mu(s) + w(s)}_{=Y(s)} + \varepsilon(s), \quad s \in \mathcal{S}$$

$\mu(\cdot)$**:** is a deterministic systematic drift $\mu(s) = \mathrm{E}(Z(s))$ modeling large scale variation;

$w(\cdot)$**:** models smaller scale variation, it is modeled as a Gaussian process

$$w(\cdot) \sim \mathrm{GP}\left(0, c(\cdot, \cdot)\right)$$

with covariance function $c : \mathcal{S} \times \mathcal{S} \to \mathbb{R}$ and mean function zero.

$\varepsilon(\cdot)$: models the additive random measurement error, it is modeled as a Gaussian process

$$\varepsilon(\cdot) \sim \mathrm{GP}\left(0, c_{\mathrm{nugget}}\left(\cdot, \cdot | \sigma^2\right)\right)$$

with nugget covariance function $c_{\mathrm{nugget}}\left(s, s' | \sigma^2\right) = \sigma^2 1_{\{0\}}\left(|s - s'|\right)$.

Assuming, $w(\cdot) \perp \varepsilon(\cdot)$, it is implied

$Y(\cdot)$: models the noisiness signal

$$Y(\cdot) \sim \mathrm{GP}\left(\mu(\cdot), c(\cdot, \cdot)\right)$$

$Z(\cdot)$: is the marginal process

$$Z(\cdot) \sim \mathrm{GP}\left(\mu(\cdot), c_Z(\cdot, \cdot)\right)$$

where $c_Z(s, s') = c(s, s') + \sigma^2 1_{\{0\}}\left(|s - s'|\right)$ assuming that $w(\cdot)$ and $\varepsilon(\cdot)$ are independent.

Note 195. The resulted hierarchical model is

(17.9)
$$\begin{cases} Z|Y \sim \mathrm{N}\left(Y, I\sigma^2\right) & \text{data model} \\ Y \sim \mathrm{N}\left(\mu(S), C(S, S)\right) & \text{spatial process model} \end{cases}$$

where $[Z]_i = Z(s_i)$, $[Y]_i = Y(s_i)$, $[\mu(S)]_i = \mu(s_i)$, and $[C(S, S)]_{i,j} = c(s_i, s_j)$.

The predictive distribution of $Y(\cdot)|Z$ under (17.9).

Note 196. Assume a vector of "unseen" sites $S_* = (s_{*,1}, ..., s_{*,q})^\top$ for any $q \in \mathbb{N}_0$. The join marginal distribution of $(Y_*, Z)^\top$ where $Z := Z(S)$ and $Y_* := Y(S_*)$ is

(17.10)
$$\begin{pmatrix} Y_* \\ Z \end{pmatrix} \sim \mathrm{N}\left(\begin{pmatrix} \mu(S_*) \\ \mu(S) \end{pmatrix}, \begin{pmatrix} C(S_*, S_*) & (C(S_*, S))^\top \\ C(S_*, S) & C(S, S) + I\sigma^2 \end{pmatrix}\right)$$

by using convenient notation $[C(S_*, S)]_{i,j} = c(s_{*,i}, s_j)$ and $[\mu(S)]_i = \mu(s_i)$.

Note 197. The (posterior) predictive distribution of $Z_*|Z$ is the conditional distribution

(17.11)
$$Y_*|Z \sim \mathrm{N}\left(\mu_1(S_*), C_1(S_*, S_*)\right)$$

where

(17.12)
$$C_1(S_*, S_*) = C(S_*, S_*) + (C(S, S_*))^\top \left(C(S, S) + I\sigma^2\right)^{-1} C(S, S_*)$$

(17.13)
$$\mu_1(S_*) = \mu(S_*) - (C(S, S_*))^\top \left(C(S, S) + I\sigma^2\right)^{-1} (\mu(S) - Z)$$

[We used Fact 192].

Created on 2024/11/04 at 11:00:38 by Georgios Karagiannis

Note 198. Since derivation of (17.11) holds for all vectors $S_* \in \mathbb{R}^q$ and all $q > 0$, (17.11) can be extended to a Gaussian Process

(17.14)
$$Y(\cdot)|Z \sim \mathrm{GP}(\mu_1(\cdot), c_1(\cdot, \cdot))$$

with

$$c_1(s, s') = c(s, s') + (C(S, s))^\top (C(S, S) + I\sigma^2)^{-1} C(S, s')$$
$$\mu_1(s) = \mu(s) - (C(S, s))^\top (C(S, S) + I\sigma^2)^{-1} (\mu(S) - Z)$$

for any $s, s' \in \mathcal{S}$. This is the predictive process for noiseless signal $Y(s)$ at any $s \in \mathcal{S}$ given Z. *[Here we used the definition of GP (Definition 17) given Note 196].*

The predictive distribution of $Z(\cdot)|Z$ under (17.9).

Note 199. It is a Gaussian Process

(17.15)
$$Z(\cdot)|Z \sim \mathrm{GP}(\mu_1(\cdot), c_1^z(\cdot, \cdot))$$

with $c_1^z(s, s') = c_1(s, s') + \sigma^2 1_{\{0\}}(|s - s'|)$. This is because $Z(\cdot) = Y(\cdot) + \varepsilon(\cdot)$ with $Y(\cdot) \perp \varepsilon(\cdot)$.

Learning $\mu(\cdot)$ and $c(\cdot, \cdot)$ via parametrization.

Note 200. Let $\mu(\cdot)$ be parametrised as $\mu(s|\beta) = \psi(s)^\top \beta$ with known basis functions $\psi(\cdot) = (\psi_1(\cdot), ..., \psi_p(\cdot))^\top$, and unknown random parameter $\beta \in \mathbb{R}^p$ following prior distribution $\beta \sim \mathrm{N}(b, B)$, $B > 0$. Let $c(\cdot, \cdot|\theta)$ be parametrised as $c(\cdot, \cdot) = c(\cdot, \cdot|\theta)$ with unknown fixed parameter $\theta \in \Theta$. Let σ^2 be an unknown fixed parameter.

Note 201. The Bayesian hierarchical model summaries to

(17.16)
$$\begin{cases} Z|Y, \sigma^2 \sim \mathrm{N}(Y, I\sigma^2) & \text{data model} \\ Y|\beta, \theta \sim \mathrm{N}(\mu(S), c(S, S)) & \text{spatial process model} \\ \beta \sim \mathrm{N}(b, B) & \text{hyper-prior model} \end{cases}$$

Note 202. The posterior of β given data Z and θ is computed via the Bayes theorem

$$\mathrm{pr}(\beta|Z, \theta, \sigma^2) \propto \mathrm{pr}(Z|\beta, \theta, \sigma^2) \mathrm{pr}(\beta)$$
$$\propto \mathrm{N}(Z|\Psi\beta, C(S, S|\theta) + I\sigma^2) \mathrm{N}(\beta|b, B)$$

and results as

(17.17)
$$\beta|Z, \theta, \sigma^2 \sim \mathrm{N}(b_n(\theta, \sigma^2), B_n(\theta, \sigma^2))$$

Created on 2024/11/04 at 11:00:38 by Georgios Karagiannis

with

$$B_n\left(\theta,\sigma^2\right) = \left(B^{-1} + \Psi^\top \left(C\left(S,S|\theta\right) + I\sigma^2\right)^{-1}\Psi\right)^{-1}$$

$$b_n\left(\theta,\sigma^2\right) = B_n\left(\theta,\sigma^2\right)\left(B^{-1}b + \Psi^\top\left(C\left(S,S|\theta\right) + I\sigma^2\right)^{-1}Z\right)$$

[For the calculations of the derivation see Exercise 18 in the Exercise sheet.]

Note 203. Given a vector of "unseen" sites $S_* = \left(s_{*,1},...,s_{*,q}\right)^\top$ for any $q \in \mathbb{N}_0$, I integrate 17.11 with respect to (17.17) i.e.

$$\mathrm{pr}\left(Y_*|Z,\theta,\sigma^2\right) = \int \mathrm{pr}\left(Y_*|Z,\beta,\theta,\sigma^2\right)\mathrm{pr}\left(\beta|Z,\theta,\sigma^2\right)\mathrm{d}\beta$$

$$= \int \mathrm{N}\left(Y_*|\mu_1\left(S_*|\beta,\theta,\sigma^2\right),C_1\left(S_*,S_*|\theta,\sigma^2\right)\right)\mathrm{N}\left(\beta|b_n\left(\theta,\sigma^2\right),B_n\left(\theta,\sigma^2\right)\right)\mathrm{d}\beta$$

where now $C\left(\cdot|\theta,\sigma^2\right)$, $C_1\left(\cdot|\theta,\sigma^2\right)$ and $\mu_1\left(\cdot|\theta,\sigma^2,\beta\right)$ are parameterized by β, θ, and σ^2. By using Fact 190

(17.18) $$Y_*|Z,\theta,\sigma^2 \sim \mathrm{N}\left(\mu_2\left(S_*|\theta,\sigma^2\right),C_2\left(S_*,S_*|\theta,\sigma^2\right)\right)$$

where

(17.19)

$$C_2\left(S_*,S_*|\theta,\sigma^2\right) = C_1\left(S_*,S_*|\theta,\sigma^2\right)$$
$$+ \left[\Psi\left(S_*\right) - \left(C\left(S,S_*|\theta\right)\right)^\top\left(C\left(S,S|\theta\right) + I\sigma^2\right)^{-1}\Psi\left(S_*\right)\right]$$
$$\times B_n\left(\theta,\sigma^2\right)\left[\Psi\left(S_*\right) - \left(C\left(S,S_*|\theta\right)\right)^\top\left(C\left(S,S|\theta\right) + I\sigma^2\right)^{-1}\Psi\left(S_*\right)\right]^\top$$

(17.20)

$$\mu_2\left(S_*|\theta,\sigma^2\right) = \Psi\left(S_*\right)b_n\left(\theta,\sigma^2\right) - \left(C\left(S,S_*|\theta\right)\right)^\top\left(C\left(S,S|\theta\right) + I\sigma^2\right)^{-1}\left(\Psi\left(S\right)b_n\left(\theta,\sigma^2\right) - Z\right)$$

Note 204. Since derivation of (17.18) holds for all vectors $S_* \in \mathbb{R}^q$ and all $q > 0$, (17.18) can be extended to a Gaussian Process

(17.21) $$Y\left(\cdot\right)|Z,\theta,\sigma^2 \sim \mathrm{GP}\left(\mu_2\left(\cdot|\theta,\sigma^2\right),c_2\left(\cdot,\cdot|\theta,\sigma^2\right)\right)$$

(17.22)

$$\mu_2\left(s|\theta,\sigma^2\right) = \psi\left(s\right)b_n\left(\theta,\sigma^2\right) - \left(C\left(s|\theta\right)\right)^\top\left(C\left(\theta\right) + I\sigma^2\right)^{-1}\left(\Psi b_n\left(\theta,\sigma^2\right) - Z\right)$$

(17.23)

$$c_2\left(s,s'|\theta,\sigma^2\right) = c_1\left(s,s'|\theta,\sigma^2\right)$$
$$+ \left[\psi\left(s\right) - \left(C\left(s|\theta\right)\right)^\top\left(C\left(\theta\right) + I\sigma^2\right)^{-1}\Psi\right]B_n\left(\theta,\sigma^2\right)\left[\psi\left(s\right) - \left(C\left(s|\theta\right)\right)^\top\left(C\left(\theta\right) + I\sigma^2\right)^{-1}\Psi\right]$$

with column vector $C\left(s|\theta\right) = \left(c\left(s,s_1\right),...,c\left(s,s_n\right)\right)^\top$, and matrix $C\left(\theta\right) := C\left(S,S|\theta\right)$.

Note 205. Estimates $\hat{\theta}$ and $\hat{\sigma}^2$ of the unknown fixed hyper-parameters θ and σ^2 are computed by maximizing the marginal likelihood of Z given θ and σ^2

(17.24) $$\text{pr}\left(Z|\theta,\sigma^2\right) = \int \text{pr}\left(Z|\beta,\theta,\sigma^2\right)\text{pr}\left(\beta\right)\text{d}\beta$$

(17.25) $$= \int \text{N}\left(Z|\Psi\beta, C\left(\theta\right)+I\sigma^2\right)\text{N}\left(\beta|b,B\right)\text{d}\beta$$

(17.26) $$= \text{N}\left(Z|\Psi b, C\left(\theta\right)+I\sigma^2+\Psi B\Psi^\top\right)$$

where $C\left(\theta\right) := C\left(S,S|\theta\right)$ [from Fact 190] by computing

$$\left(\hat{\theta},\hat{\sigma}^2\right) = \underset{\theta,\sigma^2}{\arg\min}\left(-2\log\left(\text{N}\left(Z|\Psi b, C\left(\theta\right)+I\sigma^2+\Psi B\Psi^\top\right)\right)\right)$$

The predictive distribution of $Z\left(\cdot\right)|Z$ after parameterization.

Note 206. It is a Gaussian Process

(17.27) $$Z\left(\cdot\right)|Z,\theta,\sigma^2 \sim \text{GP}\left(\mu_2\left(\cdot|\theta,\sigma^2\right), c_2^z\left(\cdot,\cdot|\theta,\sigma^2\right)\right)$$

with $c_2^z\left(s,s'|\theta,\sigma^2\right) = c_2\left(s,s'|\theta,\sigma^2\right)+\sigma^2 1_{\{0\}}\left(|s-s'|\right)$. This is because $Z\left(\cdot\right) = Y\left(\cdot\right)+\varepsilon\left(\cdot\right)$ with $Y\left(\cdot\right)\perp\varepsilon\left(\cdot\right)$.

Note 207. The estimated "Kriging predictor" results by plugging $\hat{\theta}$ and $\hat{\sigma}^2$ in (17.21)

(17.28) $$Y\left(\cdot\right)|Z,\hat{\theta},\hat{\sigma}^2 \sim \text{GP}\left(\mu_2\left(\cdot|\hat{\theta},\hat{\sigma}^2\right), c_2\left(\cdot,\cdot|\hat{\theta},\hat{\sigma}^2\right)\right)$$

Part 3. **Spatial misalignment (special topic)**

18. REGULARIZATION (AN INTRO)

Note 208. Let $\left(Z\left(s\right): s\in\mathcal{S}\right)$ be a random field with mean $\mu\left(s\right)$ at $s\in\mathcal{S}$, covariance function $c\left(s,s'\right)$ at $s,s'\in\mathcal{S}$, and semivariogram (if exists) $\gamma\left(h\right)$.

Definition 209. Let $f\left(t\right)$ be an integrable function of t such as $\int|f\left(t\right)|\text{d}t < \infty$. The regularized random field $\left(Z_f\left(s\right): s\in\mathcal{S}\right)$ is defined as

(18.1) $$Z_f\left(s\right) = \int Z\left(s+t\right)f\left(t\right)\text{d}t$$

Example 210. The mean, covariance function, and semivariogram $\left(Z_f\left(s\right): s\in\mathcal{S}\right)$ in (18.1) is

$$\mu_f\left(s\right) = \text{E}\left(Z_f\left(s\right)\right) = \int \mu\left(s+t\right)f\left(t\right)\text{d}t$$

$$c_f\left(s,s'\right) = \text{Cov}\left(Z_f\left(s\right),Z_f\left(s'\right)\right) = \int\int c\left(s+t,s'+t'\right)f\left(t\right)f\left(t'\right)\text{d}t\text{d}t'$$

Proof. See Exercise 21 in the Exercise sheet. $\qquad\square$

Note 211. Regularization is often motivated by applications such as

Prediction: In mining, one wants to predict not a single value of the process at a new site, but instead the total ore content in a block of rock.

Data collection: One cannot observe a random field at an individual site t, but can observe only the average value of the process in a small region about t. E.g., in mining, the smallest practical measurement that can be made is the average mineral content over a section of a borehole.

Image: there is often some smearing between neighboring pixels of an image.

Note 212. Here the concept is presented on simple applications of spatial misalignment and change of support

$$f(t) = \frac{1}{|B|} 1 (t \in B), \quad B \subseteq \mathcal{S}.$$

19. INTRO TO SPATIAL MISALIGNMENT

Note 213. Consider a stochastic process $(Z(s) : s \in \mathcal{S})$ where $\mathcal{S} \in \mathbb{R}^d$, $d \in \mathbb{N}_{>0}$, with $\text{Var}(s) < \infty$ for all $s \in \mathcal{S}$.

Definition 214. We define the block average $Z(B)$ with $B \subseteq \mathcal{S}$ as

(19.1)
$$Z(B) = \begin{cases} \frac{1}{|B|} \int_B Z(s) \, dx & |B| > 0 \\ \text{average} \{Z(s) : s \in B\} & |B| = 0 \end{cases}$$

where $|B| = \int 1_B(s) \, ds$.

Definition 215. The integral in (19.1) can be defined by Riemann sums. E.g. in 2D if $B = [a_1, a_2] \times [b_1, b_2]$, $a_1 < u_0 < ... < u_n < a_2$, $b_1 < v_0 < ... < v_n < b_2$, $u'_j \in [u_{j-1}, u_j]$, and $v'_i \in [v_{i-1}, v_i]$, then

(19.2)
$$\int_B Z(s) \, ds = \lim_{n \to \infty, m \to \infty} \sum_{i=1}^n \sum_{j=1}^m (v_i - v_{i-1})(u_j - u_{j-1}) Z(v'_i, u'_j)$$

Note 216. Notice that the integral in (19.1) is a linear operator, hence for $A, B \subseteq \mathcal{S}$ it is

$$\text{E}(Z(A)) = \text{E}\left(\frac{1}{|A|} \int_A Z(s) \, ds\right) = \frac{1}{|A|} \int_A \text{E}(Z(s)) \, ds$$

$$\text{Cov}(Z(A), Z(B)) = \frac{1}{|A|} \frac{1}{|B|} \int_A \int_B \text{Cov}(Z(s), Z(t)) \, ds dt$$

Note 217. A common problem is to predict the block average $Z(B)$ of a process $(Z(s) : s \in \mathcal{S})$ over a block B whose location and geometry are known and whose d-dimensional volume is $|B|$. (See Figure 21.1)
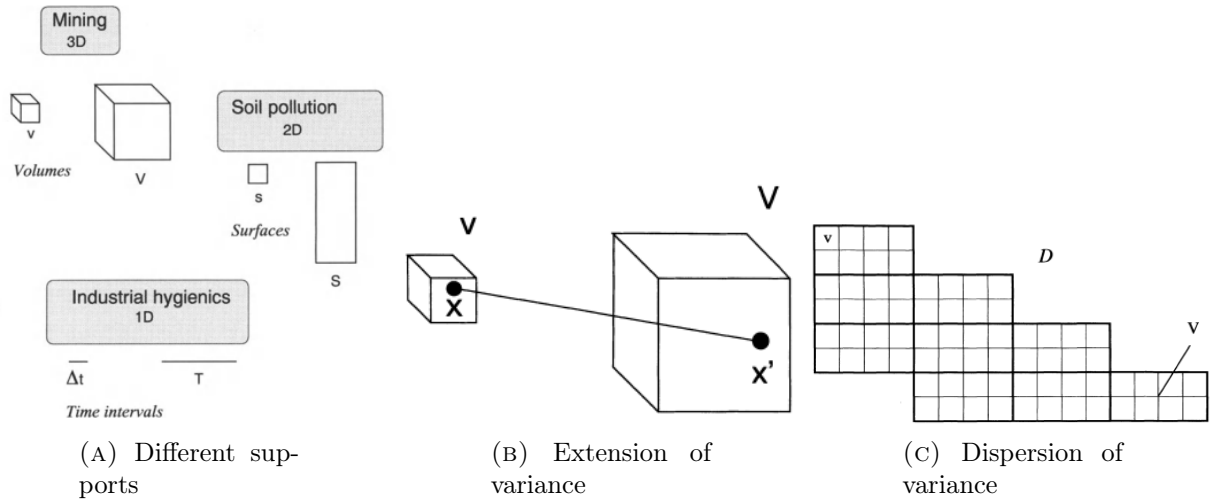
Created on 2024/11/04 at 11:00:38 by Georgios Karagiannis

(A) Different supports

(B) Extension of variance

(C) Dispersion of variance

FIGURE 19.1. Change of support

Definition 218. The support of the block average $Z(B)$ in (19.1) is B and involves the geometry, size, and spatial orientation of the line, area, or volume of the input.

Change of support problem.

Note 219. Changing the support of a variable creates a new variable related to the original one but with different statistical characteristics: mean, co-variance, dependencies, etc...

Definition 220. Change of support problem refers to making inference on block of averages whose supports are different than those of the data. $Z = (Z(B_1), ..., Z(B_n))^\top$.

20. EXTENSION AND DISPERSION VARIANCE

Note 221. With spatial variables it is necessary to take account the spatial disposal of points, surfaces or volumes for which the variance of a quantity should be computed.

Definition. Extension variance $\sigma_E^2(v, V)$ of a small volume v to a larger volume V is defined by

$$\sigma_E^2(v, V) := \mathrm{Var}(Z(v) - Z(V))$$

Definition 222. The dispersion variance of of the small identical volumes v_j partitioning a larger volume V is

$$\sigma^2(v|V) = \frac{1}{n} \sum_{j=1}^{n} \sigma_E^2(v_j, V)$$

Created on 2024/11/04 at 11:00:38 by Georgios Karagiannis

Notation 223. Let v be a small volume v and let V be a larger volume. Then we denote a semivariogram integral

$$\bar{\gamma}\left(v,V\right) = \frac{1}{|v|\,|V|} \int_{s \in v} \int_{s \in V} \gamma\left(s - s'\right) \mathrm{d}s \mathrm{d}s'$$

Proposition 224. *Let $\left(Z\left(s\right): s \in \mathcal{S}\right)$ be an intrinsic random field with semivariogram $\gamma\left(\cdot\right)$.*

(1) *The extension variance $\sigma_E^2\left(v,V\right)$ of a small volume v to a larger volume V is*

$$\sigma_E^2\left(v,V\right) = 2\bar{\gamma}\left(v,V\right) - \bar{\gamma}\left(v,v\right) - \bar{\gamma}\left(V,V\right)$$

(2) *Suppose a large volume V is partitioned into n smaller units $\{v_j\}_{j=1}^n$ of equal size and geometry. The dispersion variance of each unit v_j is*

(20.1) $$\sigma^2\left(v|V\right) = \bar{\gamma}\left(V,V\right) - \bar{\gamma}\left(v,v\right)$$

Proof. For part 1, see Exercise 22 in the Exercise sheet. For part 2, see Exercise 24 in the Exercise sheet. $\square$

Note 225. The following result (Krige's relationship) resembles to "ANOVA".

Proposition 226. *[Krige's relationship] Consider that the domain S is partitioned into volumes V which are partitioned into smaller volumes v; i.e. $v \subseteq V \subseteq S$. Then the relation between the three supports is*

(20.2) $$\sigma^2\left(v|\mathcal{S}\right) = \sigma^2\left(v|V\right) + \sigma^2\left(V|\mathcal{S}\right)$$

Proof. (Sketch of the proof) (20.1) becomes

$$\sigma^2\left(v|\mathcal{S}\right) = \bar{\gamma}\left(\mathcal{S},\mathcal{S}\right) - \bar{\gamma}\left(v,v\right)$$

similar

$$\sigma^2\left(v|V\right) = \bar{\gamma}\left(V,V\right) - \bar{\gamma}\left(v,v\right)$$

so (20.1) gives

(20.3) $$\sigma^2\left(v|\mathcal{S}\right) - \sigma^2\left(v|V\right) = \sigma^2\left(V|\mathcal{S}\right)$$

$\square$

Change of support effect.

Note 227. Consider the case that the domain S is partitioned into volumes V which are partitioned into smaller volumes v. Assume there are available samples at "point" locations s each of them lies to the center of one of the smaller volumes v. Making the assumption that the sampled value at each point location s is extended to each area of influence v implies that the distribution of average values of the blocks is the same as the distribution
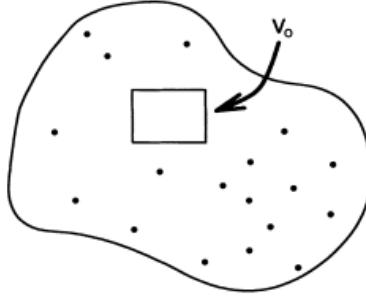
FIGURE 21.1. Block Kriging cartoon

of the values at the sample points. However from (20.3), we see that this is not true; in fact the distribution of the values for a support v is narrower than the distribution of point values because the variance $\sigma^2(s|v)$ of the points in v generally is not negligible; i.e. $\sigma^2(s|V) - \sigma^2(v|V) = \sigma^2(s|v) > 0$.

Change of support: affine model.

Note 228. Consider a stationary process $Z(s)$ for $s \in \mathcal{S}$, and consider a block process $Z_v(s)$ on a block v. The affine model assumes that the standardized point variable $Z(s)$ follows the same distribution as the standardized block variable $Z(v)$.

Example 229. An example of the use of affine models is the Gaussian process case, where $Z(s) \sim \mathrm{N}(\mu, \sigma^2)$ and $Z(v) \sim \mathrm{N}(\mu, \sigma_v^2)$, –same mean but different variances– it is

$$\frac{Z(s) - \mu}{\sqrt{\sigma^2}} \overset{\text{distr.}}{\sim} \frac{Z(v) - \mu}{\sqrt{\sigma_v^2}} \sim \mathrm{N}(0, 1)$$

which implies the relation

$$Z(v) \overset{\text{distr.}}{\sim} \mu + \sqrt{\frac{\sigma_v^2}{\sigma^2}} (Z(s) - \mu) \sim \mathrm{N}(\mu, \sigma_v^2)$$

21. BLOCK KRIGING

Note 230. Block Kriging (BK) aims to predict a block value $Z(v_0)$ at block v_0 instead of at a point value s_0; see Figure 21.1. It can be used within the framework of Universal, Ordinary, Simple, and Bayesian Kriging cases we saw in Section 16.1.

Note 231. Assume we want the estimate a block value $Z(v_0)$ at block v_0 with some volume $|v_0|$ given that my data $\{(s_i, Z_i)\}_{i=1}^n$ are realizations $Z_i = Z(s_i)$ at point values s_i (Figure 21.1).

21.1. **(Classical) Universal Block Kriging.**

Created on 2024/11/04 at 11:00:38 by Georgios Karagiannis

Note 232. Here, we present the Block Kriging (BK) in the (Classical) Universal Kriging framework (Section 16.1). We will refer to the UK in Section 16.1 as point-to-point UK.

Note 233. Consider that the statistical model is the stochastic process $(Z(s) : s \in \mathcal{S})$ with

$$(21.1) \qquad\qquad Z(s) = \mu(s) + \delta(s)$$

Assume

$$\mu(s) = \sum_{j=0}^{p} \psi_j(s)\beta_j = (\psi(s))^\top \beta$$

with vector of unknown coefficients $\beta = (\beta_0, ..., \beta_p)^\top$ and vector of known basis functions $\psi(s) = (\psi_0(s), ..., \psi_p(s))^\top$. Assume $\delta(s)$ is a zero mean process. Assume $\delta(s)$ is an intrinsic stationary process with a semi-variogram $\gamma(\cdot)$ –as in UK in Section 16.1; intrinsic stationarity is not a necessary assumption if one can estimate the covariance function directly.

Note 234. The Block UK predictor $Z_{\mathrm{BK}}(v_0)$ of $Z(v_0)$ at block v_0 with support $|v_0| > 0$ has the following linear form weighted by a set of tunable unknown weights

$$(21.2) \qquad\qquad Z_{\mathrm{BK}}^*(v_0) = w_{n+1} + \sum_{i=1}^{n} w_i Z(s_i) = w_{n+1} + w^\top Z$$

where $Z = (Z_1, ..., Z_n)^\top$ and $w = (w_1, ..., w_n)^\top$.

Note 235. Following the steps in (point-to-point) UK (Note 155), consideration of $\psi_0(\cdot) = 1$ unbiasness implies conditions

$$(21.3) \qquad \text{ASSUMPTION:} \qquad\qquad \Psi_0 = \sum_{i=1}^{n} w_i \psi(s_i) \Leftrightarrow \Psi_0 = w^\top \Psi$$

$$(21.4) \qquad \text{ASSUMPTION:} \qquad\qquad w_{n+1} = 0$$

$$(21.5) \qquad \text{ASSUMPTION:} \qquad\qquad \sum_{i=1}^{n} w_i = 1 \Leftrightarrow w^\top \underline{1} = 1$$

where $[\Psi_0]_j = \psi_j(v_0)$, and $\psi_j(v_0) = \frac{1}{|v_0|}\int \psi_j(s)\,\mathrm{d}s$ for $j = 0, ..., p$.

Note 236. Following the steps in (point-to-point) UK (Note 159), I get

$$(21.6) \qquad \mathrm{MSE}(Z_{\mathrm{BK}}(v_0)) - \sum_{i=1}^{n} w_i \sum_{j=1}^{n} w_j \gamma(s_i - s_j) + 2\sum_{i=1}^{n} w_i \bar{\gamma}(s_i, v_0)$$

$$= -w^\top \mathbf{\Gamma} w + 2w^\top \bar{\boldsymbol{\gamma}}_0 - \bar{\boldsymbol{\gamma}}_{00}$$

where $\bar{\boldsymbol{\gamma}}_{00} = \bar{\gamma}(v_0, v_0) \neq 0$, $\bar{\boldsymbol{\gamma}}_0 = (\bar{\gamma}(s_1, v_0), ..., \bar{\gamma}(s_n, v_0))^\top$, and $\bar{\gamma}(s_i, v_0)$ be the average variogram of each sample point with the block of interest. This is the same as that of point-to-point UK in (16.8) where the point $\gamma(s_i, s_0)$ is substituted by the integral $\bar{\gamma}(s_i, v_0)$.

Note 237. The Block Universal Kriging equations then are

(21.7)
$$\begin{cases} 0 = -2\mathbf{\Gamma} w_{\mathrm{BK}} + 2\bar{\boldsymbol{\gamma}}_0 - \Psi \lambda_{\mathrm{UK}} \\ \Psi_0 = w_{\mathrm{BK}}^\top \Psi \end{cases}$$

which essentially produce the same weights as the point-to-point Universal Kriging but averaged out in the block

(21.8)
$$w_{\mathrm{BK}} = \mathbf{\Gamma}^{-1} \left(\bar{\boldsymbol{\gamma}}_0 - \Psi \left(\Psi^\top \Gamma^{-1} \Psi \right)^{-1} \left(\Psi^\top \Gamma^{-1} \bar{\boldsymbol{\gamma}}_0 - \Psi_0^\top \right) \right)$$

(21.9)
$$\lambda_{\mathrm{BK}} = 2 \left(\Psi^\top \Gamma^{-1} \Psi \right)^{-1} \left(\Psi^\top \Gamma^{-1} \boldsymbol{\gamma}_0 - \Psi_0^\top \right)$$

Note 238. Hence the UK predictor $Z_{\mathrm{BK}}(s_0)$ at s_0 is

(21.10)
$$Z_{\mathrm{BK}}(s_0) = \left(\bar{\boldsymbol{\gamma}}_0 - \Psi \left(\Psi^\top \Gamma^{-1} \Psi \right)^{-1} \left(\Psi^\top \Gamma^{-1} \bar{\boldsymbol{\gamma}}_0 - \Psi_0^\top \right) \right)^\top \mathbf{\Gamma}^{-1} Z$$

with standard error (by substituting (21.7) in (21.6))

(21.11)
$$\sigma_{\mathrm{BK}}(v_0) = \sqrt{-\bar{\gamma}_{00} - w_{\mathrm{BK}}^\top \mathbf{\Gamma} w_{\mathrm{BK}} + 2 w_{\mathrm{BK}}^\top \bar{\boldsymbol{\gamma}}_0}$$

Note 239. Block Kriging as a concept can be implemented even when s_i are not points but have some volume $|s_i| > 0$. Then we call the case as aggregation if $|s_i| < |v_0|$, or disaggregation if $|s_i| > |v_0|$.

21.2. **Bayesian Block Kriging.**

Note 240. If $Z(s)$ is a Gaussian process defined on points $s \in \mathcal{S}$, then the block average $Z(v)$ with $v \subseteq \mathcal{S}$ is a Gaussian process as well. This is because integration (or averaging) in (19.1) is a linear operation as seen in (19.2), and linear combinations of Gaussians is Gaussian as well.

Note 241. Bayesian Block Kriging predictive distribution and moments are derived in the same way as the "point to point" Bayesian Kriging in Section 17. Let $(Z(s) : s \in \mathcal{S})$ be a GRF, and $\{(Z(s_i), s_i) ; s_i \in \mathcal{S}\}$ be observables. I want to compute the Bayesian Block Kriging predictive distribution of $Z(v_0)$ at unseen volume/unit $v_0 \subseteq \mathcal{S}$. Main steps involve.

(1) compute the joint distribution in (17.10) i.e.

$$\begin{pmatrix} Z(v_0) \\ Z(\{s_i\}) \end{pmatrix} \sim \mathrm{N} \left(\begin{pmatrix} \mu(v_0) \\ \mu(\{s_i\}) \end{pmatrix}, \begin{pmatrix} c(v_0, v_0) & c(v_0, \{s_i\}) \\ (c(v_0, \{s_i\}))^\top & c(\{s_i\}, \{s_i\}) \end{pmatrix} \right)$$

with

$$\mu(v_0) = \frac{1}{|v_0|} \int_{x \in v_0} \mu(x) \, \mathrm{d}x$$

$$c(v_0, s_i) = \frac{1}{|v_0|} \int_{x \in v_0} c(x, s_i) \, \mathrm{d}x$$

$$c(v_0, v_0') = \frac{1}{|v_0| \, |v_0'|} \int_{x \in v_0} \int_{y \in v_0'} c(x, y) \, \mathrm{d}x \mathrm{d}y$$

(2) compute the the predictive distribution as the conditional Normal distribution $\mathrm{pr}\left(Z(v_0) \, | Z\right)$ (Note 196), and

(3) recognize the corresponding Gaussian process as in Note 198.

The derivation is identical to that Section 17.2.