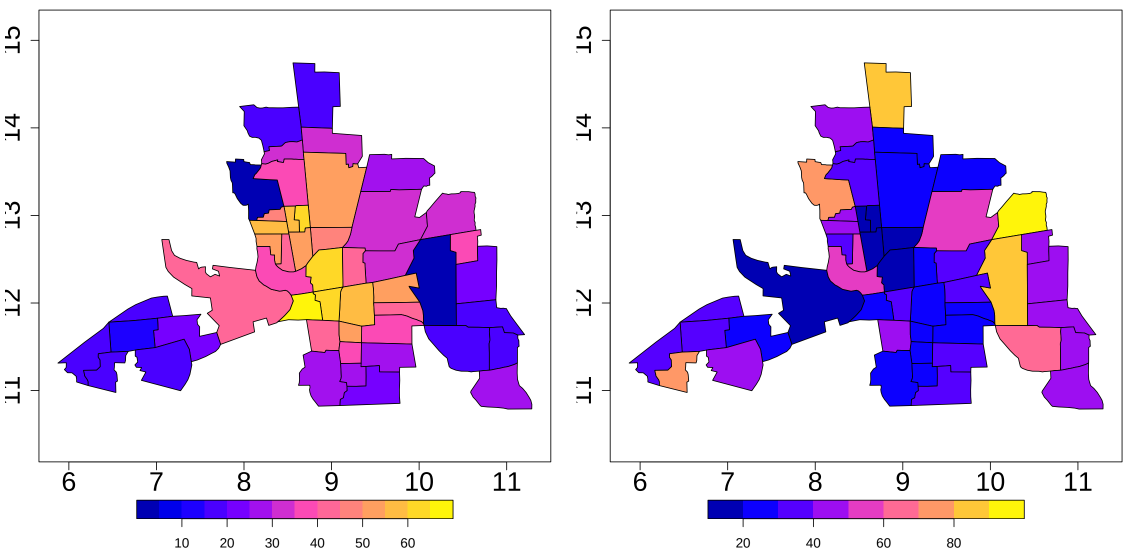
Spatio-temporal statistics (MATH4341) Michaelmas term, 2024

Exercise sheet

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

Part 1. Types of spatial data

($\star$)(Columbus Columbus OH data set) Figure 1a shows the Property crime (number per thousand households) in 49 districts in Columbus in 1980, as well as the average value of the house in USD. Figure 1b presents the corresponding average house value. This is the R dataset `columbus{spdep}`. Interest may lie to find whether high rates of crime are clusterd in a particular areas, and if yes, perhaps what is the association of it with the value of the houses in the area. To which principal spatial statistical are would you associate this problem?



(A) Property crime (number per thousand households)

(B) Average value of the house in USD

FIGURE 1. Columbus Columbus OH spatial analysis dataset

Exercise 1. ($\star$)(Soil chemistry properties data set.) It contains measurements of various chemical properties of soil samples collected at different locations in a field. These properties include: the acidity or alkalinity of the soil (PH), the salt concentration in the soil (Salinity), and others. It is the R dataset `soil250{geoR}`. Figure 2 presents the locations these measurements are taken. The data (measurements) are in fixed locations at a regular grid of points. The domain scientist would be interested in the nutrient levels and pH to assess soil fertility and make recommendations for agricultural practices. The statistician could (i.) estimate/predict values of soil properties at unsampled locations based on measurements at sampled locations; and (ii.) assess the spatial variability of soil properties (nutrient levels and pH) to identify regions with high or low variability. To which principal spatial statistical are would you associate this problem?

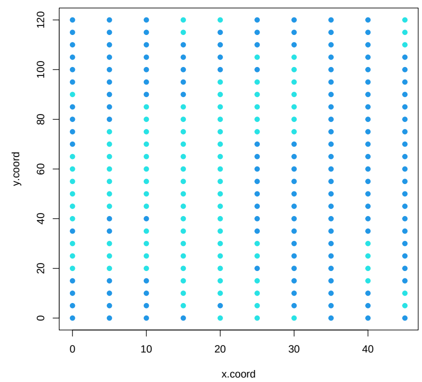


FIGURE 2. Soil chemistry data set

Exercise 2. ($\star$)(Wolfcamp-aquifer data) Figure 3 presents locations and levels (in feet above sea level) of piezometric head for the aquifer; they are obtained by drilling a narrow pipe into the aquifer and letting the water find its own level in the pipe. After rigorous screening of unsuitable wells, 85 remained. There is interest to find where the radionuclide contamination would flow from the site in Deaf Smith County, Texas. Beneath Deaf Smith County is a deep brine aquifer known as the Wolfcamp aquifer, a potential pathway for any radionuclides leaking from the repository. The predicted direction of flow can be used to determine locations of downgradient and upgradient wells for a groundwater monitoring system. A first direction in analyzing this spatial data set is to draw a map of a predicted surface based on the (irregularly located) 85 data. To which principal spatial statistical are would you associate this problem?

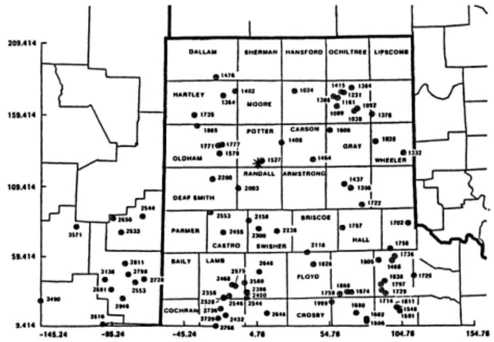


FIGURE 3. Wolfcamp-aquifer data. Piezometric-head levels (feet above sea level) vs coordinates.

Exercise 3. ($\star$)(Swiss rainfall data) Figure 4 presents the locations of the 100 locations in Switzerland as dots whose size and color indicates the amount of the corresponding rainfall measurements (in 10th of mm) taken on May 8, 1986. This is the R data set `SIC{geoR}`. Observation sites are irregularly spaced, and fixed. A scientific objective may be to analyzing rainfall patterns with purpose to optimize crop planting and irrigation schedules. A statistician is able to estimate rainfall values at unsampled locations based on available measurements, create maps that represent the spatial distribution of rainfall, or quantify the uncertainty associated with rainfall estimates and predictions, which are important for risk assessment and decision-making. To which principal spatial statistical are would you associate this problem?

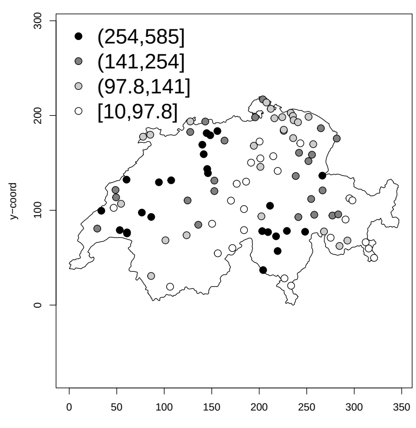


FIGURE 4. Swiss rainfall data

Exercise 4. ($\star$)(Air pollution in Piemonte.) Figure 5 presents the average PM10 ($\mu g/m^3$) concentration during October 2005–March 2006 for the 24 monitoring stations in the Piemonte region (Northern Italy). The data (measurements) are at fixed locations at irregular grid points. PM10 is one of the most troublesome pollutants in the area. Environmental agencies need models to predict PM10 at unmonitored sites in order to assess PM10 concentration over an entire region. A geostatistician can build a model which is satisfactory in terms of goodness of fit, interpretability, parsimony, prediction capability and computational costs with purpose to build reliable PM10 concentration maps, equipped with the corresponding uncertainty measure.To which principal spatial statistical are would you associate this problem?

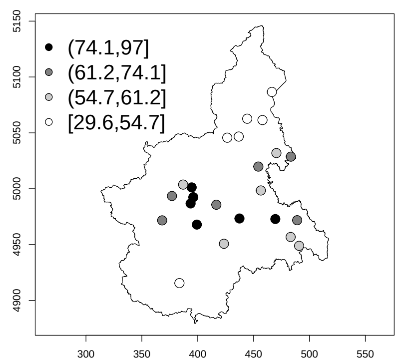


FIGURE 5. (Air pollution data) Average PM10 ($\mu g/m^3$) concentration during October 2005–March 2006 for the 24 monitoring stations in the Piemonte

---

Exercise 5. ($\star$)(Scallop abundance data) The scallop data is based on a 1990 survey cruise in the Atlantic continental shelf off Long Island, New York, U.S.A. They are available from R as `scallop {SemiPar}`. Figure 6 presents 148 locations (degrees of longitude & latitude) in the Atlantic waters off the coasts of New Jersey and Long Island New York as coordinates and the size of scallop catch at the corresponding location as the dot size. The sites are at fixed locations within an irregular grid of points. Sustainable scallop abundance is critical for the long-term economic viability of the fishing industry. A healthy and stable scallop population supports a consistent source of income for fishermen and related businesses.

(1) To which principal spatial statistical are would you associate this problem?
(2) Can you suggest a Bayesian hierarchical model for this? Justify your suggestion.

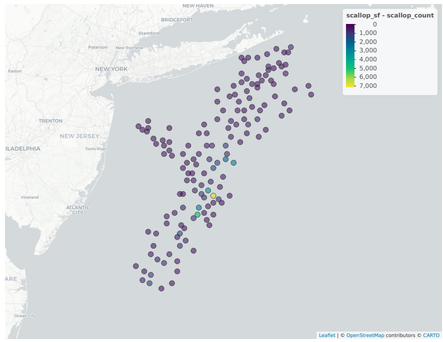


Figure 6. Scallop abundance data

Exercise 6. (⋆)An example of spatio-temporal data where aerial spatial data are time referenced is given in Figure 7 which shows a spatio-temporal dataset representing the population of the counties of Ohio, USA, from 1968 to 1988. The dataset is available from the SpatialEpiApp R package. Interest lies in not only how the population is only distributed over the spatial domain but also how it evolves during the time. To which principal spatial statistical are would you associate this problem?

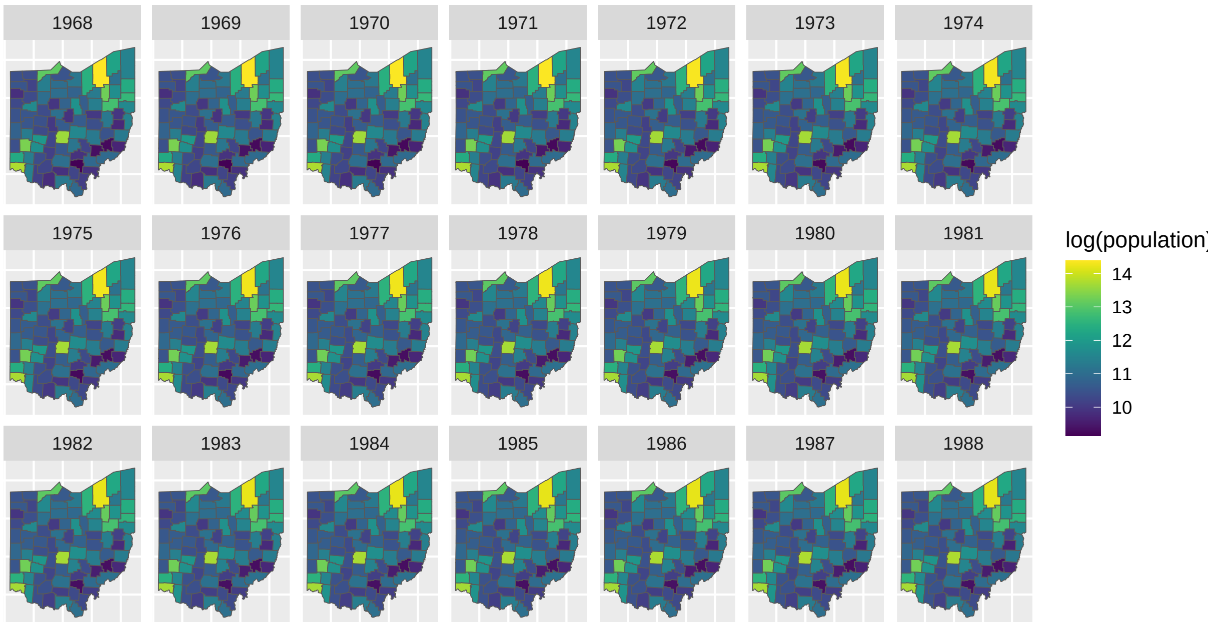


FIGURE 7. Population of the counties of Ohio, USA, from 1968 to 1988.

Part 2. **Point referenced data / Geostatistics**

Exercise 7. ($\star$)If $c : \mathbb{R}^d \to \mathbb{R}$ is the covariogram of a weakly stationary random field $Z = (Z_s)_{s\in\mathbb{R}^d}$ then $c(\cdot)$ is semi-positive definite; i.e. for all $n \in \mathbb{N}$, $a \in \mathbb{R}^n$, and $\{s_1, ..., s_n\} \subseteq S$

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

Exercise 8. ($\star$) Show that if $c_1(\cdot,\cdot)$ and $c_2(\cdot,\cdot)$ are are covariance functions (are non-negative definite) then so are $c_3(\cdot,\cdot) = bc_1(\cdot,\cdot) + dc_2(\cdot,\cdot)$ with $b, d \geq 0$ and $c_4(\cdot,\cdot) = c_1(\cdot,\cdot)\, c_2(\cdot,\cdot)$.

Exercise 9. ($\star$) Consider the Gaussian c.f. $c(h) = \sigma^2 \exp\left(-\beta \|h\|_2^2\right)$ for $\sigma^2, \beta > 0$ and $h \in \mathbb{R}^d$. Compute the spectral density from Bochner's theorem

---

Exercise 10. ($\star$) Consider the Exponential c.f. $c(h) = \sigma^2 \exp\left(-\beta \|h\|_1^1\right)$ for $\sigma^2, \beta > 0$ and $h \in \mathbb{R}^d$. Compute the spectral density from Bochner's theorem

---

(Given as Formative assessment 1)

Exercise 11. ($\star$) Let $Z = \left(Z(s) : s \in \mathbb{R}^d\right)$ be an intrinsic random field with $\mathrm{E}(Z(s) - Z(t)) = 0$ and let $\gamma : \mathbb{R}^d \to \mathbb{R}$ be its semivariogram.

(1) Let $a \in \mathbb{R}^n$ be a vector of constants. Consider sites $\left\{s_1, ..., s_n \subseteq \mathbb{R}^d\right\}$ Show that

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

where $c_I(s,t) = \gamma(s - s_0) + \gamma(t - s_0) - \gamma(s - t)$ at some additional $s_0 \in \mathbb{R}^d$.

(2) Show that 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

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

(Given as Formative assessment 1)

Exercise 12. ($\star$) Consider the zero-mean random field $Z = \left(Z(s) : s \in \mathbb{R}^d\right)$ with covariogram function given by

$$c(h) = \begin{cases} \xi^2 (1 + \rho \|h\|) \exp(-\rho \|h\|), & h > 0 \\ \nu^2 + \xi^2, & h = 0 \end{cases}$$

(1) Compute the semivariogram for the random field $\left(Z(s) : s \in \mathbb{R}^d\right)$

(2) What are the nugget, sill and partial sill for this covariance model? Justify your answer.

(3) Would the slightly altered covariance function defined below be a good model for spatial data for $\phi > 0$? Justify your answer.

$$c(h) = \begin{cases} \xi^2 (1 + \rho \|h\|) \exp(-\rho \|h\|) + \phi, & h > 0 \\ \nu^2 + \xi^2 + \phi, & h = 0 \end{cases}$$

(Given as Formative assessment 2)

Exercise 13. ($\star$) Consider we the geostatistical model $(Z(s); s \in \mathcal{S})$ with

$$Z(s) = \mu(s) + w(s) + \varepsilon(s)$$

where $w(s)$ is a weakly stationary process with mean zero and covariogram $c_w(h; \sigma^2, \phi) = \sigma^2 \exp\left(-\frac{1}{\phi}\|h\|\right)$, $\mu(s; \beta)$ is a deterministic function

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

with unknown coefficients $\beta = (\beta_0, ..., \beta_p)^\top$ and known basis functions $\psi(s) = (\psi_0(s), ..., \psi_p(s))^\top$, $\varepsilon(s)$ is a nugget effect process whose covariogram has sill τ^2, and assume that $w(s)$ and $\varepsilon(s)$ are independent Gaussian Processes.

(1) Write down the formula of the covariogram $c(h; (\sigma^2, \phi, \tau))$ of (Z_s).

(2) Consider a re-parametrization $\theta = (\sigma^2, \phi, \xi)$ where $\xi^2 = \frac{\tau^2}{\sigma^2}$ is called signal to noise ratio. Assume there is available a dataset $\{(s_i, Z_i)\}_{i=1}^n$ where $, Z_i := Z(s_i)$ is a realization of $(Z(s); s \in \mathcal{S})$ at site s_i.

   (a) Let Ψ be a matrix with $[\Psi]_{i,j} = \psi_j(s_i)$. Let D be a matrix such as $[D]_{i,j} = \|s_i - s_j\|$. Consider that you can use convenient notation such as $\exp(D)$ meaning $[\exp(D)]_{i,j} = \exp(D_{i,j})$. Write down the covariance matrix $C(\theta)$ of $Z = (Z_1, ..., Z_n)^\top$ as a function of D and θ.

   (b) Write down the log likelihood function $\log(L(Z; \theta))$ of $Z = (Z_1, ..., Z_n)^\top$ given $\theta = (\sigma^2, \phi, \xi)$.

(3) Let $r(\cdot)$ (called correlogram) such as $c(\cdot) = \sigma^2 r(\cdot)$. Assume that (ϕ, ξ) as known constants.

   (a) Compute the likelihood equations[1] w.r.t. (β, σ^2), and for given (ϕ, ξ).

   (b) Compute the MLE $\hat{\beta}_{(\phi,\xi)}$ of β as a function of (ϕ, ξ)

   (c) Compute the MLE $\hat{\sigma}^2_{(\phi,\xi)}$ of σ^2 as a function of (ϕ, ξ).

   (d) Compute the unbiased estimator of $\tilde{\sigma}^2$ of σ^2.

      Hint: Consider the fitted values $e = (e_1, ..., e_n)^\top$ as $e = [I - H] Z$ where $H = (\Psi^\top R^{-1} \Psi)^{-1} \Psi^\top R^{-1}$, and write $\hat{\sigma}^2_{(\phi,\xi)}$ w.r.t. e.

      Hint: It is given that $\mathrm{E}(Z^\top A Z) = \mathrm{E}(Z)^\top A \mathrm{E}(Z)^\top + \mathrm{tr}(A \mathrm{Var}(Z))$ when $Z \sim$ Normal

[1] that is, the gradient of the log-likelihood

(4) Compute the so-called log "profiled likelihood" $\log\left(L\left(Z;(\phi,\xi)\right)\right)$ resulting as

$$L\left(Z;(\phi,\xi)\right)=L\left(Z;\beta=\hat{\beta}_{(\phi,\xi)},\sigma^2=\hat{\sigma}^2_{\left(\hat{\beta}_{(\phi,\xi)},\phi,\xi\right)},\phi,\xi\right)$$

by replacing the β with $\hat{\beta}_{(\phi,\xi)}$ and σ^2 with $\hat{\sigma}^2_{\left(\hat{\beta}_{(\phi,\xi)},\phi,\xi\right)}$ in the actual likelihood $L\left(Z;\beta,\theta=\left(\sigma^2,\phi,\xi\right)\right)$. Describe how you would compute suitable values $\left(\hat{\phi},\hat{\xi}\right)$ for the MLE of (ϕ,ξ)

Exercise 14. ($\star$) Let $(Z(s):s\in\mathcal{S})$ be a specified statistical model. Assume that $(Z(s):s\in\mathcal{S})$ is weakly stationary with unknown constant mean $\mu=\mathrm{E}(Z(s))$ and known covariogram $c(\cdot)$. Assume there is available a dataset $\{(s_i,Z_i:=Z(s_i))\}_{i=1}^n$ and assume they are realizations of $(Z_s)_{s\in\mathcal{S}}$. Assume that the matrix C such as $[C]_{i,j}=c(\|s_i-s_j\|)$ has an inverse. Consider the "Kriging" estimator μ_{KM} of μ as the BLUE (Best Linear Unbiased Estimator)

$$\mu_{\mathrm{KM}}=\sum_{i=1}^n w_iZ(s_i)=w^\top Z,$$

for some unknown $\{w_i\}$ that we need to learn.

(1) Find sufficient conditions on $w=(w_1,...,w_n)$ so that the Kriging estimator μ_{KM} to be unbiased.
(2) Assume C is invertable. Compute the MSE of μ_{KM} as a function of $w=(w_1,...,w_n)$ and C
(3) Derive the Kriging estimator μ_{KM} of μ as a function of C
(4) Derive the Kriging standard error as $\sigma_{\mathrm{KM}}=\sqrt{\mathrm{E}\left(\mu_{\mathrm{KM}}-\mu\right)^2}$ as a function of C

(Given as Formative assessment 2)

Exercise 15. ($\star$) Let $(Z(s);s\in\mathcal{S})$ be a specified statistical model. Assume that process $(Z(s);s\in\mathcal{S})$ has known mean $\mu(s)=\mathrm{E}(Z(s))$ and known covariance function $c(\cdot,\cdot)$. Assume there is available a dataset $\{(s_i,Z_i:=Z(s_i))\}_{i=1}^n$. Assume that the matrix C such as $[C]_{i,j}=c(s_i,s_j)$ has an inverse. Consider the "Kriging" estimator μ_{SK} Consider the "Kriging" estimator $Z_{\mathrm{SK}}(s_0)$ of $Z(s_0)$ at an unseen spatial location s_0 as the BLUE (Best Linear Unbiased Estimator)

$$Z_{\mathrm{SK}}(s_0)=w_{n+1}+\sum_{i=1}^n w_iZ(s_i)=w_{n+1}+w^\top Z,$$

for some unknown $\{w_i\}$ that we need to learn, and $Z=(Z_1,...,Z_n)^\top$. Let $w=(w_1,...,w_n)^\top$.

(1) Find sufficient conditions on $w=(w_1,...,w_n)^\top$ so that the Kriging estimator $Z_{\mathrm{SK}}(s_0)$ to be unbiased.
(2) Derive the MSE of $Z_{\mathrm{SK}}(s_0)$ as

$$\mathrm{E}\left(Z_{\mathrm{SK}}(s_0)-Z(s_0)\right)^2=w^\top Cw+c(s_0,s_0)-2w^\top C_0$$

where C_0 is a vector such as $[C_0]_i=c(s_0,s_i)$.

(3) Derive the Kriging estimator of $Z(s_0)$ as

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

where $\mu(s_{1:n})$ is a vector such as $[\mu(s_{1:n})]_i = \mu(s_i)$.

(4) Compute the Kriging standard error $\sigma_{\text{SK}} = \sqrt{\text{E}(Z_{\text{SK}}(s_0) - Z(s_0))^2}$.

Exercise 16. ($\star$) Assume a spatial model

$$Z(s) = \mu + \delta(s), \ s \in \mathcal{S} \tag{1}$$

with unknown mean $\mu \in \mathbb{R}$. Assume a set of n observed realizations $Z_i := Z(s_i)$ of (1) at sites s_i for $i = 1, ..., n$. Assume that $Z(s)$ is a weak stationary stochastic process with known covariogram $c(\cdot)$. Derive the formula for the Ordinary Kriging predictor $Z_0 := Z(s_0)$ at spatial location s_0 and its kriging variance as function of the covariogram $c(h)$ and not the semi-variogram.

Exercise 17. ($\star$) Let $(Z(s); s \in \mathcal{S})$ be a specified statistical model. Assume that $(Z(s); s \in \mathcal{S})$ is an intrinsic stationary process with unknown constant mean $\mu = \text{E}(Z(s))$ and known semi-variogram $\gamma(\cdot)$. Assume there is available a dataset $\{(s_i, Z_i := Z(s_i))\}_{i=1}^n$. Consider the "Kriging" estimator $Z_{\text{OK}}(s_0)$ of $Z(s_0)$ at any unseen spatial location s_0 as the BLUE (Best Linear Unbiased Estimator)

$$Z_{\text{OK}}(s_0) = w_{n+1} + \sum_{i=1}^n w_i Z(s_i) = w_{n+1} + w^\top Z$$

for some unknown $\{w_i\}$ that we need to learn, and $Z = (Z_1, ..., Z_n)^\top$. Let $w = (w_1, ..., w_n)^\top$.

(1) Find sufficient conditions on $w = (w_1, ..., w_n)$ so that the Kriging estimator $Z_{\text{OK}}(s_0)$ to be unbiased.

(2) Derive the MSE of $Z_{\text{OK}}(s_0)$ as

$$\text{E}(Z_{\text{OK}}(s_0) - Z(s_0))^2 = -w^\top \mathbf{\Gamma} w + 2w^\top \boldsymbol{\gamma}_0$$

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

(3) Assume Γ is invertable matrix. Derive the Kriging estimator of $Z(s_0)$ as

$$Z_{\text{OK}}(s_0) = \mathbf{\Gamma}^{-1} \left(\boldsymbol{\gamma}_0 + \frac{1 - 1^\top \mathbf{\Gamma}^{-1} \boldsymbol{\gamma}_0}{1^\top \mathbf{\Gamma}^{-1} 1} 1 \right) Z$$

(4) Derive the Kriging standard error of $Z_{\text{OK}}(s_0)$ as

$$\sigma_{\text{SK}} = \sqrt{\boldsymbol{\gamma}_0 \mathbf{\Gamma}^{-1} \boldsymbol{\gamma}_0 - \frac{\left(1 - 1^\top \mathbf{\Gamma}^{-1} \boldsymbol{\gamma}_0\right)^2}{1^\top \mathbf{\Gamma}^{-1} 1}}$$

Exercise 18. ($\star$)

Inventory of useful formulas.

[Normal distr. conditioning] Let $x_1 \in \mathbb{R}^{d_1}$, and $x_2 \in \mathbb{R}^{d_2}$. If

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

then it is

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

where

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

Consider the Bayesian Kriging from your lecture notes:

$$Z(s) = Y(s) + \varepsilon(s), \quad s \in \mathcal{S}$$

where

$$\varepsilon(\cdot) \sim \mathrm{GP}\left(0, c_\varepsilon(\cdot, \cdot|\tau)\right)$$

with $c_\varepsilon(s, s'|\tau) = \tau^2 1_{\{0\}}(\|s - s'\|)$ and

$$Y(\cdot)|\beta, \theta \sim \mathrm{GP}\left(\mu(\cdot|\beta), c_Y\left(\cdot, \cdot|\sigma^2, \phi\right)\right)$$

with mean function $\mu(\cdot|\beta)$ (to be specified later) labeled by unknown parameter β and covariance function $c_Y\left(\cdot, \cdot|\sigma^2, \phi\right)$.

Assume there is available a dataset $\{(s_i, Z_i)\}_{i=1}^n$ where $Z_i = Z(s_i)$ is a realization of a stochastic process (Z_s).

(1) Write the hierarchical spatial model $Z(\cdot)|Y(\cdot), \beta, \varphi$ and $Y(\cdot)|\beta, \varphi$ where $\varphi = \left(\sigma^2, \phi, \tau\right)^\top$.

(2) Write the marginal process $Z(\cdot)|\beta, \varphi$ where $\varphi = \left(\sigma^2, \phi, \tau\right)^\top$, its mean function denoted as $\mu(\cdot|\cdot)$, and its covariance function denoted as $c(\cdot|\cdot)$.

Hint:: Let Y and X be independent random variables with $X \sim \mathrm{N}(\mu_X, \Sigma_X)$, $Y \sim \mathrm{N}(\mu_Y, \Sigma_Y)$. Let A and B be fixed matrices. Let c be a fixed vector. Then

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

(3) Compute the predictive process $Z(\cdot)|Z, \beta, \varphi$ as

$$Z(\cdot)|Z, \beta, \varphi \sim \mathrm{GP}\left(\mu_1(\cdot|\beta, \varphi), c_1(\cdot, \cdot|\varphi)\right)$$

with

$$\begin{aligned} c_1\left(s, s'|\varphi\right) =& c(s, s|\varphi) + \left(C(S, s|\varphi)\right)^\top \left(C(S, S|\varphi)\right)^{-1} C\left(S, s'|\varphi\right) \\ \mu_1(s|\beta, \varphi) =& \mu(s|\beta) - \left(C(S, s|\varphi)\right)^\top \left(C(S, S|\varphi)\right)^{-1}\left(\mu(S|\beta) - Z\right) \end{aligned}$$

Hint: See the Conditional Normal formula above.

(4) Assume $\mu(s|\beta) = \psi(s)^\top \beta$. Consider a conjugate prior $\beta \sim \mathrm{N}(b, B)$ on β where $B > 0$.

(a) Write down the Bayesian statistical model involving layers $[Z|\beta, \varphi]$, and $[\beta|\varphi]$.

(b) Compute the posterior distribution as

$$\beta|Z,\varphi \sim \mathrm{N}\left(b_n\left(\varphi\right), B_n\left(\varphi\right)\right)$$

with

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

where $C\left(S,S|\varphi\right)$ is a matrix with $\left[C\left(S,S|\varphi\right)\right]_{i,j} = c\left(s_i,s_j|\varphi\right)$.

Hint: Use the following identity

$$(y-\Phi\beta)^\top\Sigma^{-1}(y-\Phi\beta) + (\beta-\mu)^\top V^{-1}(\beta-\mu) = \left(\beta-\mu^*\right)^\top \left(V^*\right)^{-1}\left(\beta-\mu^*\right) + S^*;$$
$$V^* = \left(V^{-1} + \Phi^\top\Sigma^{-1}\Phi\right)^{-1}; \qquad \mu^* = V^*\left(V^{-1}\mu + \Phi^\top\Sigma^{-1}y\right)$$
$$S^* = \mu^\top V^{-1}\mu - \left(\mu^*\right)^\top\left(V^*\right)^{-1}\left(\mu^*\right) + y^\top\Sigma^{-1}y;$$

(c) Compute the (posterior) predictive process $Z\left(\cdot\right)|Z,\varphi$ given the data Z and given the parameters φ as

$$Z\left(\cdot\right)|Z,\varphi \sim \mathrm{GP}\left(\mu_2\left(\cdot|\varphi\right), c_2\left(\cdot,\cdot|\varphi\right)\right)$$

with

$$\mu_2\left(s|\varphi\right) = \left(\psi\left(s\right) - \Psi^\top C^{-1}C\left(s\right)\right)^\top \left(B^{-1} + \Psi^\top C^{-1}\Psi\right)^{-1} B^{-1}b$$
$$+ \left[\left(C\left(s\right)\right)^\top + \left(\psi\left(s\right) - \Psi^\top C^{-1}C\left(s\right)\right)^\top \left(B^{-1} + \Psi^\top C^{-1}\Psi\right)^{-1}\Psi\right] C^{-1}Z$$
$$c_2\left(s,s'|\varphi\right) = c\left(s,s'|\varphi\right) - \left(C\left(s\right)\right)^\top C^{-1}C\left(s'\right)$$
$$+ \left(\psi\left(s\right) - \Psi^\top C^{-1}C\left(s\right)\right)^\top \left(B^{-1} + \Psi^\top C^{-1}\Psi\right)^{-1}\left(\psi\left(s'\right) - \Psi^\top C^{-1}C\left(s'\right)\right)$$

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

(d) Compute the marginal likelihood $\mathrm{pr}\left(Z|\varphi\right)$ in the form

$$\mathrm{pr}\left(Z|\sigma^2,\varphi\right) = \mathrm{N}\left(Z|\Psi b, \left(C^{-1} - C^{-1}\Psi\left(B^{-1} + \Psi^\top B^{-1}\Psi\right)^{-1}\Psi^\top C^{-1}\right)^{-1}\right)$$

where Ψ is a matrix with $\left[\Psi\right]_{i,j} = \psi_j\left(s_i\right)$, and R is a matrix with $\left[C\right]_{i,j} = c\left(s_i,s_j|\varphi\right)$.

Hint-2:: It is

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

Hint 3:: [Woodbury matrix identity]

$$\left(A+UCV\right)^{-1} = A^{-1} - A^{-1}U\left(C^{-1} + VA^{-1}U\right)^{-1}VA^{-1}$$

(5) Consider non-informative prior $\mathrm{pr}\left(\beta\right) \propto 1$ for β by specifying $b \to 0$ and letting $B^{-1} \to 0$. Argue whether such a prior can be used. Recompute the (asymptotic) quantities $\mathrm{pr}\left(Z|\varphi\right)$, $\left[Z\left(\cdot\right)|Z,\varphi\right]$ under this new prior in the limit.

Exercise 19. ($\star$) Consider the Bayesian Kriging in the Lecture notes, in particular the "Gaussian process regression".

The Bayesian hierarchical model summaries to

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

The estimates $\hat{\theta}$ and $\hat{\sigma}^2$ of the unknown fixed hyper-parameters θ and σ^2 are said to be computed as

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

Show that if I consider noiseless observations (which can be achieved by $\sigma^2 \to 0$) and uniform hyper-prior for β i.e. $\mathrm{pr}\left(\beta\right) \propto 1$ (which can be achieved a $b \to 0$ and $B^{-1} \to 0$) then I get the following results.

(1) The asymptotic posterior of β under the assumptions of noiseless data and uniform prior of β is proper.

Hint 1: The posterior of β under the Bayesian model (2) is

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

with

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

Hint 2: (Woodbury matrix identity)

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

(2) The asymptotic predictive random field $Y\left(\cdot\right)|Z,\theta$ is

$$Y\left(\cdot\right)|Z,\theta \sim \mathrm{GP}\left(\tilde{\mu}_2\left(\cdot|\theta\right), \tilde{c}_2\left(\cdot,\cdot|\theta\right)\right)$$

$$\tilde{\mu}_2\left(s|\theta\right) = \left[\left(\psi\left(s\right) - \left(C\left(s|\theta\right)\right)^\top C^{-1}\Psi\right)\left(\Psi^\top C\left(\theta\right)^{-1}\Psi\right)^{-1}\Psi^\top + \left(C\left(s|\theta\right)\right)^\top\right]C^{-1}Z$$

$$\begin{aligned} \tilde{c}_2\left(s,s'|\theta\right) = &\left[\psi\left(s\right) - \left(C\left(s|\theta\right)\right)^\top C\left(\theta\right)^{-1}\Psi\right]\left(\Psi^\top C\left(\theta\right)^{-1}\Psi\right)^{-1}\left[\psi\left(s'\right) - \left(C\left(s'|\theta\right)\right)^\top C\left(\theta\right)^{-1}\Psi\right]^\top \\ &+ c\left(s,s'|\theta\right) + \left(C\left(s|\theta\right)\right)^\top C\left(\theta\right)^{-1}C\left(s'|\theta\right) \end{aligned}$$

aka the same as those of Universal Kriging. Here $C\left(\theta\right) := C\left(S,S|\theta\right)$ and $C\left(s|\theta\right) = \left(c\left(s,s_1|\theta\right),...,c\left(s,s_n|\theta\right)\right)^\top$.

Hint: The predictive random field $Y(\cdot)|Z,\theta,\sigma^2$ under the Bayesian model (2) is

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

$$\begin{aligned}
\mu_2\left(s|\theta,\sigma^2\right) =& \psi(s)\, b_n\left(\theta,\sigma^2\right) - \left(C(s|\theta)\right)^\top \left(C(\theta)+I\sigma^2\right)^{-1}\left(\Psi b_n\left(\theta,\sigma^2\right) - Z\right)\\
c_2\left(s,s'|\theta,\sigma^2\right) =& c_1\left(s,s'|\theta,\sigma^2\right)\\
&+\left[\psi(s) - \left(C(s|\theta)\right)^\top \left(C(\theta)+I\sigma^2\right)^{-1}\Psi\right] B_n\left(\theta,\sigma^2\right)\left[\psi(s) - \left(C(s|\theta)\right)^\top \left(C(\theta)+I\sigma^2\right)^{-1}\Psi\right]\\
c_1\left(s,s'|\theta,\sigma^2\right) =& c\left(s,s'|\theta\right) + \left(C(S,s|\theta)\right)^\top\left(C(S,S|\theta)+I\sigma^2\right)^{-1} C\left(S,s'|\theta\right)
\end{aligned}$$

(3) the estimate $\hat{\theta}$ of the unknown fixed hyper-parameter θ is computed as

$$\begin{aligned}
\hat{\theta} = \arg\min_{\theta,\sigma^2} \Bigg(& -\frac{n}{4}\log\left(\left|C(\theta)^{-1} - C^{-1}\Psi\left(\Psi^\top C(\theta)^{-1}\Psi\right)^{-1}\Psi^\top C(\theta)^{-1}\right|\right)\\
& + Z^\top\left[C(\theta)^{-1} - C(\theta)^{-1}\Psi\left(\Psi^\top C(\theta)^{-1}\Psi\right)^{-1}\Psi^\top C(\theta)^{-1}\right] Z\Bigg)
\end{aligned}$$

where $C(\theta) := C(S,S|\theta)$.

Hint: *The* estimates $\hat{\theta}$ and $\hat{\sigma}^2$ of the unknown fixed hyper-parameters θ and σ^2 under the Bayesian model (2) is are

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

Exercise 20. ($\star$) Assume that $(Z(s): s\in\mathcal{S})$ is a a random field with semivariogram $\gamma(\cdot)$.

(1) Then

$$\mathrm{Var}\left(\int Z(s+t)\, g(t)\right) = -\frac{1}{2}\int g(s)\, g(t)\, \gamma(s-t)\,\mathrm{d}s\mathrm{d}t$$

for any $g(t)$ be an integrable function of t such as

$$\int g(t)\,\mathrm{d}t = 0$$

(2) Semivariogram $\gamma(\cdot)$ is conditionally negative definite (c.n.d.): if for all $n\in\mathbb{N}$, for all function g over $\mathcal{S}$ such that $\int g(t)\,\mathrm{d}t = 0$

$$\int g(s)\, g(t)\, \gamma(s-t)\,\mathrm{d}s\mathrm{d}t \leq 0,$$

Exercise 21. ($\star$) Let $(Z(s): s\in\mathcal{S})$ be a random field with mean $\mu(s)$ at $s\in\mathcal{S}$, and covariance function $c(s,s')$ at $s,s'\in\mathcal{S}$.. Let $f(t)$ be an integrable function of t such as $\int|f(t)|\,\mathrm{d}t<\infty$. Let regularized random field $(Z_f(s): s\in\mathcal{S})$ defined as

$$Z_f(s) = \int Z(s+t)\, f(t)\,\mathrm{d}t$$

(1) Show that the mean of $Z_f(s)$ is

$$\begin{aligned}\mu_f(s) =& \mathrm{E}(Z_f(s)) \\ =& \int \mu(s+t) f(t)\,\mathrm{d}t\end{aligned}$$

(2) Show that the covariance function of of $Z_f(s)$ is

$$\begin{aligned}c_f(s,s') =& \mathrm{Cov}\left(Z_f(s), Z_f(s')\right) \\ =& \int\int c\left(s+t, s'+t'\right) f(t) f\left(t'\right) \mathrm{d}t\mathrm{d}t\end{aligned}$$

(3) Assume that $(Z(s) : s \in \mathcal{S})$ is a a random field with semivariogram $\gamma(h)$. Show that for any $g(t)$ be an integrable function of t such as

$$\int g(t)\,\mathrm{d}t = 0$$

it is

$$\mathrm{Var}\left(\int Z(s+t) g(t)\right) = -\frac{1}{2}\int g(s) g(t) \gamma(s-t)\,\mathrm{d}s\mathrm{d}t$$

(4) Assume that $(Z(s) : s \in \mathcal{S})$ is a a random field semivariogram with semivariogram $\gamma(h)$. Show that

$$\begin{aligned}\gamma_f(h) =& \frac{1}{2}\mathrm{Var}(Z_f(h) - Z_f(0)) \\ =& \int\int f(s) f(t-h)\gamma(s-t)\,\mathrm{d}s\mathrm{d}t - \int\int f(s) f(t) \gamma(s-t)\,\mathrm{d}s\mathrm{d}t\end{aligned}$$

Hint:: Assume that $(Z(s) : s \in \mathcal{S})$ is a a random field with semivariogram $\gamma(h)$. Then

$$\mathrm{Var}\left(\int Z(s+t) g(t)\right) = -\frac{1}{2}\int g(s) g(t) \gamma(s-t)\,\mathrm{d}s\mathrm{d}t$$

for any $g(t)$ be an integrable function of t such as

$$\int g(t)\,\mathrm{d}t = 0$$

Exercise 22. ($\star$) Show that the extension variance $\sigma_E^2(v, V)$ of a small volume v to a larger volume V is obtained by

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

where

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

Exercise 23. ($\star$) Suppose a large volume V is partitioned into n smaller units v of equal size. Show that the dispersion variance $\sigma^2(v|V) = \frac{1}{n}\sum_{j=1}^{n} \sigma_E^2(v_j, V)$ can be written in term of variogram integrals

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

as

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

(Given as Formative assessment 3)

Exercise 24. ($\star$) Consider a statistical model which is a stochastic process $(Z_s)_{s\in\mathbb{R}}$ (so s has dimension 1), where $Z(\cdot) \sim \mathrm{GP}(\mu(\cdot), c(\cdot,\cdot))$ with mean function $\mu(s) = 1$ and covariance function $c(s,t) = \exp\left(-(s-t)^2\right)$ for any $s \in \mathbb{R}$ and $t \in \mathbb{R}$. Assume there is available a dataset $\{(Z_i, s_i)\}_{i=1}^n$ where $Z_i = Z(s_i)$ and $s_i \in \mathbb{R}$ are point sites.

(1) Compute the length $|v|$ of the block $v = [a,b] \subset \mathbb{R}$.
(2) Compute the block mean $\mu(v)$ for some block $v = [a,b] \subset \mathbb{R}$ and point $s \in \mathbb{R}$.
(3) Compute the block covariance function $c(v,s)$ for some block $v = [a,b] \subset \mathbb{R}$ and point $s \in \mathbb{R}$.
(4) Compute the block covariance function $c(v,v')$ for some blocks $v = [a,b] \subset \mathbb{R}$ and $v' = [a',b'] \subset \mathbb{R}$.
(5) Denote $Z = (Z_1, ..., Z_n)^\top$, and $S = \{s_1, ..., s_n\}$. Let $v = [a,b] \subset \mathbb{R}$ and $v' = [a',b'] \subset \mathbb{R}$ be two intervals. Compute the joint distribution of $(Z(v), Z(v'), Z)^\top$ as a function of $c(\cdot,\cdot)$, S, v, v', Z, and $\mu(\cdot)$. What is the name of the distribution and what are the parameter functions defining it?
(6) (Bayesian Kriging) Compute the predictive stochastic process $[Z(v)|Z]$ at blocks $v = [a,b] \subset \mathbb{R}$ with $|v| > 0$.

Hint-1:: Let $x_1 \in \mathbb{R}^{d_1}$, and $x_2 \in \mathbb{R}^{d_2}$. If

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

then it is

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

where

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

Hint-2: You can use that $\int \mathrm{erf}(x)\,\mathrm{d}x = x\mathrm{erf}(x) + \frac{\exp(-x^2)}{\sqrt{\pi}} + \text{const}$, when $\mathrm{erf}(x) = \frac{2}{\sqrt{\pi}}\int_0^x \exp\left(-t^2\right)\mathrm{d}t$.

(Given as Formative assessment 3)

Exercise 25. ($\star$) Assume we wish to estimate the average value in a domain V

$$Z_V = \frac{1}{|V|}\int_V Z(s)\,\mathrm{d}s$$

with the average of n sample points $\{s_i; i = 1, ..., n\}$.

$$\hat{Z} = \frac{1}{n} \sum_{i=1}^{n} Z(s_i)$$

Show that the estimation variance(or else extension variance)

$$\operatorname{Var}\left(\hat{Z} - Z_V\right) = -\frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \gamma(s_i - s_j) + \frac{1}{n|V|} \sum_{i=1}^{n} \int_V \gamma(s_i - x)\,\mathrm{d}x - \frac{1}{|V|^2} \int_{x \in B} \int_{y \in B} \gamma(x - y)\,\mathrm{d}x\mathrm{d}y$$

Hint:: Consider as known that

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

Exercise 26. ($\star$) We are interested in the mean value of some parameter $Z(s)$, say temperature, over a one-hour period segment $[0, L]$. In particular we wish to find out which of these two estimates is better: the temperature at half the hour

$$\hat{Z}_1 = Z\left(\frac{L}{2}\right)$$

or

$$\hat{Z}_2 = \frac{1}{2}(Z(0) + Z(L))$$

the average of two consecutive measurements on the hour?

Assume that $(Z(s) : s \in \mathcal{S})$ is intrinsic random field with semivariogram $\gamma(\cdot)$. Let

$$\chi(h) = \frac{1}{h} \int_0^h \gamma(u)\,\mathrm{d}u,\ h > 0$$
$$\xi(h) = \frac{1}{h^2} \int_0^h \int_0^h \gamma(x - y)\,\mathrm{d}x\mathrm{d}y,\ h > 0$$

(1) Show that the estimation variances (or else extension variances) of $\hat{Z}_1$ and $\hat{Z}_2$ to domain segment $[0, L]$ are

$$\sigma_1^2 = \operatorname{Var}\left(\hat{Z}_1 - Z_V\right) = 2\chi\left(\frac{L}{2}\right) - \xi(L)$$
$$\sigma_2^2 = \operatorname{Var}\left(\hat{Z}_2 - Z_V\right) = 2\chi(L) - \xi(L) - \frac{\gamma(L)}{2}$$

(2) Assume semivariogram $\gamma(h) = b|h|^a$. Show that

$$\chi(h) = \frac{bh^a}{a+1}, \quad \xi(h) = \frac{2bh^a}{(a+1)(a+2)}$$

(3) Assume the power semivariogram $\gamma(h) = |h|^a$. Show that

$$\sigma_1^2 = \frac{2L^a}{(a+1)(a+2)} \frac{1}{2^a} (a + 2 - 2^a)$$
$$\sigma_2^2 = \frac{2L^a}{(a+1)(a+2)} \frac{1}{4} \left(2 + a - a^2\right)$$

Which estimate is preferable for $a = 0$, and $a = 1$,

Hint: If $Z_V = \frac{1}{|V|}\int_V Z(s)\,\mathrm{d}s$ is average value in a domain V and if $\hat{Z} = \frac{1}{n}\sum_{i=1}^{n} Z(s_i)$ is average of n sample points $\{s_i; i = 1, ..., n\}$ then the the estimation variance(or else extension variance) is

$$\mathrm{Var}\left(\hat{Z} - Z_V\right) = -\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\gamma(s_i - s_j) + \frac{1}{n|V|}\sum_{i=1}^{n}\int_V \gamma(s_i - x)\,\mathrm{d}x - \frac{1}{|V|^2}\int_{x\in B}\int_{y\in B}\gamma(x-y)\,\mathrm{d}x\mathrm{d}y$$

Exercise 27. ($\star$) Consider

$$Z(s) = \sum_{p=1}^{k} a_p w_p(s)$$

where $\left\{w_p^{(u)}(s)\right\}$ are intrinsic random fields with

$$\mathrm{E}(w_p(s)) = 0$$
$$\gamma_{p,q}(h) = 0, \text{ for } p \neq q$$

$p = 1, ..., k$ and $q = 1, ..., k$. Assume $\{a_p\}$ is a set of known constants. Let $\gamma(h)$ be the variogram function of $Z(s)$. Show that

$$\gamma(h) = \sum_{p=1}^{k} a_p^2 \gamma_{p,p}(h)$$

Exercise 28. ($\star$) Consider a set of random fields $\left\{\left(Z_j^{(u)}(s) : s \in \mathcal{S}\right); j = 1, ..., k; u = 1, ..., k\right\}$ with

$$Z_j^{(u)}(s) = \sum_{p=1}^{k} a_{j,p}^{(u)} w_p^{(u)}(s),$$

where $\left\{w_p^{(u)}(s)\right\}$ are intrinsic random fields and $\left\{a_{j,p}^{(u)}\right\}$ are known constants. Let $\tilde{\gamma}_{i,j}^{(u)}(h)$ be the cross variogram function of $Z_i^{(u)}(s)$ and $Z_j^{(u)}(s)$ for $u = 1, ..., k$.

(1) Write the definition of the the cross variogram function $\tilde{\gamma}_{i,j}^{(u)}(h)$ of $Z_i^{(u)}(s)$ and $Z_j^{(u)}(s)$ for $u = 1, ..., k$

(2) Assume that

$$\mathrm{E}\left(w_p^{(u)}(s)\right) = 0$$

$$\mathrm{Cov}\left(w_p^{(u)}(s), w_q^{(v)}(s+h)\right) = \begin{cases} \gamma_{p,q}^{(u)}(h), & u = v \\ 0 & u \neq v \end{cases}$$

$u = 1, ..., k$, $p = 1, ..., k$ and $q = 1, ..., k$. Show that

$$\tilde{\gamma}_{i,j}^{(u)}(h) = \sum_{p=1}^{k} a_{i,p}^{(u)} \sum_{q=1}^{k} a_{j,q}^{(u)} \gamma_{p,q}^{(u)}(h)$$

(3) Assume that

$$\mathrm{E}\left(w_p^{(u)}\left(s\right)\right) = 0$$

$$\mathrm{Cov}\left(w_p^{(u)}\left(s\right), w_q^{(v)}\left(s+h\right)\right) = \begin{cases} \gamma^{(u)}\left(h\right), & u = v \,\mathrm{and}\, p = q \\ 0 & u \neq v \,\mathrm{or}\, p \neq q \end{cases}$$

$u = 1, ..., k$. Show that

$$\tilde{\gamma}_{i,j}^{(u)}\left(h\right) = \sum_{p=1}^{k} a_{i,p}^{(u)} \sum_{q=1}^{k} a_{j,q}^{(u)} \gamma^{(u)}\left(h\right)$$

and hence

$$\tilde{\Gamma}^{(u)}\left(h\right) = B^{(u)}\gamma^{(u)}\left(h\right)$$

where $B^{(u)} = A^{(u)}\left(A^{(u)}\right)^{\top}$, $u = 1, ..., k$.

(4) Consider the assumptions in previous part. Let $\{(Z_j\left(s\right) : s \in \mathcal{S}); j = 1, ..., k\}$ be a set of random fields on $s \in \mathcal{S}$. Let

$$Z_j\left(s\right) = \mu_j\left(s\right) + \sum_{u=0}^{m} Z_j^{(u)}\left(s\right)$$

and let

$$Z\left(s\right) = \mu\left(s\right) + \sum_{u=0}^{m} A^{(u)} w^{(u)}\left(s\right)$$

Show that the cross variogram matrix of $\left(Z\left(s\right); s \in \mathcal{S}\right)$ is

$$\Gamma\left(h\right) = \sum_{u=0}^{m} B^{(u)}\gamma^{(u)}\left(h\right)$$

where $B^{(u)} = A^{(u)}\left(A^{(u)}\right)^{\top}$, $u = 1, ..., k$

Part 3. Aerial unit data / spatial data on lattices