20

(The University Of Choice)

# MASINDE MULIRO UNIVERSITY OF SCIENCE AND TECHNOLOGY (MMUST)

UNIVERSITY EXAMINATIONS

2022/2023 ACADEMIC YEAR, SECOND SEMESTER

SECOND YEAR MAIN EXAMINATIONS

FOR THE DEGREE OF

BACHELOR OF SCIENCE IN DISASTER PREPAREDNESS AND ENGINEERING MANAGEMENT

COURSE CODE: DPG 206
COURSE TITLE: SPATIAL STATISTICS

DATE: THURSDAY, 17TH APRIL, 2023 TIME: ~~8.00 AM – 10.00 AM~~ 3–5 PM

**Instruction to the candidates:**
*Answer question ONE (COMPULSORY) and any other TWO questions*
*Time: 2 hours*

This paper consists of 3 printed pages. Please turn over.

**SECTION I: Answer ALL the questions in this section**

**QUESTION ONE - 30 MARKS (COMPULSORY)**

(a) Differentiate, giving an example in each case, between geostatistical, areal and point data **[5 mks]**

(b) Consider the spatial process $\{Z(\mathbf{s}) : \mathbf{s} \in D\}$ with zero mean and the following semi-variogram model

$$\gamma(h) = \begin{cases} \tau^2\sigma^2 h, & \text{if } h > 0 \\ 0, & \text{if } h = 0 \end{cases}$$

(i) Define what it means for a process to be weakly stationary. **[2 mks]**

(ii) By determining the sill (variance) of thus process from the semi-variogram, determine if the process is weakly stationary. **[3 mks]**

(c) Suppose that $\{Z(\mathbf{s}) : \mathbf{s} \in D\}$ is a weakly stationary geostatistical process. Fix a constant $\tau \in \mathbb{R}$, and let $G$ denote the set of locations $\mathbf{s}$ for which $Z(\mathbf{s}) > \tau$.

(i) Letting $Y(\mathbf{s}) = 1$ for each $\mathbf{s} \in G$, explain why $\{Y(\mathbf{s}) : \mathbf{s} \in D\}$ is a point process. **[2 mks]**

(ii) Letting $W(\mathbf{s}) = Z(\mathbf{s}) - \tau$ for each $\mathbf{s} \in G$, explain why $\{W(\mathbf{s}) : \mathbf{s} \in D\}$ is a marked point process. **[2 mks]**

(iii) Explain why we could be interested in studying processes such as $Y(\mathbf{s})$ and $W(\mathbf{s})$. **[2 mks]**

(d) Explain the goals of geostatistical data. **[5 mks]**

(e) Explain the main steps in exploratory data analysis. **[4 mks]**

(f) Consider a homogeneous Poisson process $Z(A) \sim \text{Poisson}(\lambda|A)$, where $|A|$ is the size of the set $A$ and and $\lambda$ is the constant first order intensity function. Now suppose a Bayesian approach to inference is adopted and the prior distribution for $\lambda$ is given by $\lambda \sim \text{Gamma}(a,\ b)$ for some constants $(a,\ b)$. Compute the posterior distribution for $\lambda$ given the data $Z(A)$, that is, what distribution is $f(\lambda|Z(A))$? **[5 mks]**

**SECTION II: Answer any TWO questions from this section**

**QUESTION TWO - 20 MARKS**

(a) Suppose that $\{Z(\mathbf{s}) : \mathbf{s} \in D\}$ is a stationary isotropic geostatistical process with zero mean and covariance function given by

$$C(h) = \begin{cases} \sigma^2 \exp\left(-\dfrac{h}{\phi}\right), & \text{if } h > 0 \\ \sigma^2 + \tau^2, & h = 0 \end{cases}$$

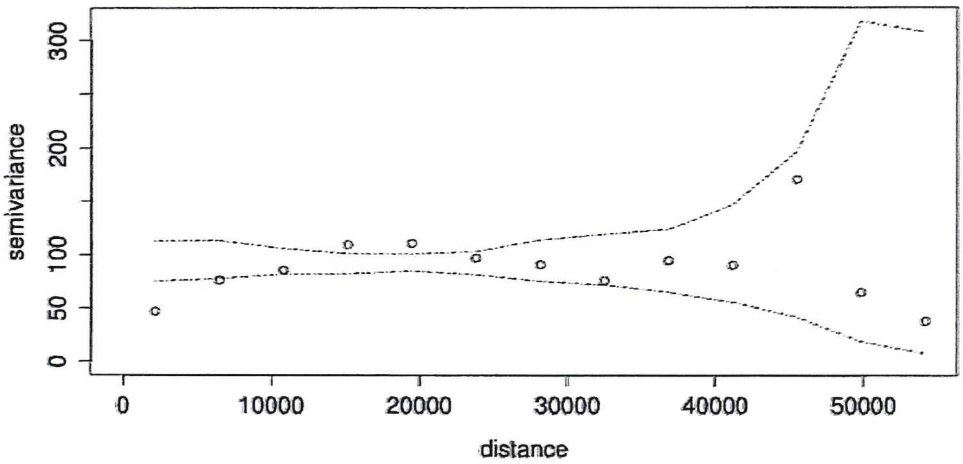
(i) Compute $C(3)$ for $\sigma = 2$ and $\tau = 5$ when (i) $\phi = 0.5$ and (ii) $\phi = 2$. Comment on the change of correlation range with the change in $\phi$. **[5 mks]**

(ii) Giving reasons for your answer, what are the range, sill, nugget, and partial sill for this covariance model? **[3 mks]**

(iii) In terms of the covariance function, derive an expression for the semi-variogram. **[2 mks]**

(iv) Sketch the shapes of covariance and the variogram functions for the model above. **[4 mks]**

(b) A binned empirical semi-variogram was plotted of a data set together with 95% Monte Carlo envelops, and ythe plot is as shown below.




(i) Describe two disadvantages with the estimation of the binned empirical semi-variogram. **[2 mks]**

(ii) Estimate the nugget, sill and range of the process. **[2 mks]**

(iii) What do you conclude about the presence of spatial autocorrelation in these data? **[2 mks]**

## QUESTION THREE - 20 MARKS

(a) Suppose that $\{Z(\mathbf{s}) : \mathbf{s} \in D\}$ for $D \subset \mathbb{R}^2$ is a stationary Gaussian process with zero mean and an isotropic Gaussian covariance function with parameters $\tau^2 = 0$, $\sigma^2 > 0$ and $\phi > 0$.

(i) Write down the log likelihood function for the data $\mathbf{z} = (z(\mathbf{s}_1), \ldots z(\mathbf{s}_n))$ **[3 mks]**

(ii) Derive the maximum likelihood estimator of $\sigma^2$ in terms of $\phi$. **[5 mks]**

(iii) Using this estimator of $\sigma^2$, write down the profile likelihood only in terms of $\phi$. Using your expression, explain why there is no closed form expression for maximum likelihood estimator of $\phi$. **[5 mks]**

(b) Let $\{Z(\mathbf{s}) : \mathbf{s} \in D\}$ for $D \subset \mathbb{R}^2$ be a stationary process with mean $\mu$ and an isotropic exponential covariance function with parameters $\tau^2 > 0$, $\sigma^2 > 0$ and $\phi > 0$. Suppose we observe data at two locations a distance one unit apart. Calculate the BLUP at location $\mathbf{s}_0$ halfway between the two observed locations. **[7 mks]**

## QUESTION FOUR - 20 MARKS

(a) Suppose that $\{X(\mathbf{s}) : \mathbf{s} \in D\}$ is a weakly stationary process with mean $\mu_X$ and covariance function $C_X(.)$ and $\{Y(\mathbf{s}) : \mathbf{s} \in D\}$ is a second weakly stationary process with mean $\mu_Y$ and covariance function $C_Y(.)$. Further let $\text{Cov}(X(\mathbf{s}), Y(\mathbf{s}+\mathbf{h})) = \omega$. Then consider

$$Z(\mathbf{s}) = \frac{1}{2}(X(\mathbf{s}) + Y(\mathbf{s})).$$

(i) Derive the mean and covariance functions and determine if $Z(\mathbf{s})$ is weakly stationary. **[5 mks]**

(ii) Calculate the covariance function and semi-variogram of $Z(\mathbf{s})$ if both $(X(\mathbf{s}),\ Y(\mathbf{s}))$ have mean zero and exponential covariance functions with common parameters $(\sigma^2,\ \tau^2,\ \phi)$. What covariance model is this? **[5 mks]**

(b) Consider a spatial domain with 4 regions ordered as $[A \mid B \mid C \mid D]$ with data values $Z(A) = 1$, $Z(B) = 2$, $Z(C) = 3$ and $Z(D) = 4$. Then consider the following two neighbourhood matrices:

$$\mathbf{W}_1 = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}, \quad \mathbf{W}_2 = \begin{pmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{pmatrix}.$$

Compute Moran's I for both neighbourhood matrices, and describe why the estimate using $\mathbf{W}_1$ is larger than the estimate using $\mathbf{W}_2$. **[5 mks]**

(c) Consider random variables $(Y_1,\ Y_2)$ that have the following conditional distributions:

$$Y_1|Y_2 \sim N(\alpha_0 + \alpha_1 Y_2, \sigma^2), \quad Y_2|Y_1 \sim N(\beta_0 + \beta_1 Y_1^3, \sigma^2).$$

Using the fact that $\mathbb{E}(X) = \mathbb{E}_Y(\mathbb{E}_X(X|Y))$, calculate $\mathbb{E}(Y_1)$ and $\mathbb{E}(Y_2)$ and define the conditions that must hold for these conditional distributions to be complete with each other. **[5 mks]**

## QUESTION FIVE - 20 MARKS

(a) For a domain $D \subset \mathbb{R}^2$, suppose $X$ is an inhomogeneous Poisson process with first order intensity function $\lambda_X(\mathbf{s})$, that is independent of $Y$, an inhomogeneous Poisson process with intensity function $\lambda_Y(\mathbf{s})$. The superposition of $X$ and $Y$ is defined to be

$$Z = X \cup Y,$$

the set of all points contained in $X$ or $Y$.

(i) Show that for any $\mathbf{s} \in D$ $\lambda_Z(\mathbf{s}) = \lambda_X(\mathbf{s}) + \lambda_Y(\mathbf{s})$. **[5 mks]**

(ii) Show that $Z$ is an inhomogeneous Poisson Process with intensity $\lambda_Z(\mathbf{s})$. **[5 mks]**

(b) For a domain $D \subset \mathbb{R}^2$, let $Z$ denote a homogeneous Poisson process with first order intensity function $\lambda$.

(i) Write down the properties that characterize this Poisson process. **[4 mks]**

(ii) Is $Z$ a stationary and isotropic spatial point process? Explain your reasoning. **[3 mks]**

(iii) Derive Ripley's K function for Z. **[3 mks]**