SOME EXAM DATE
SOME EXAM TIME
EXAMINATION FOR THE DEGREES OF XXXX

STATISTICS
Spatial Statistics 4H

"Hand calculators with simple basic functions (log, exp, square root, etc.) may be used in examinations. No calculator which can store or display text or graphics may be used, and any student found using such will be reported to the Clerk of Senate".

NOTE: Candidates should attempt all questions.

1. (i) Consider the geostatistical process $\{Z(\mathbf{s})|\mathbf{s} \in \mathcal{D}\}$.

   (a) Define what it means for $Z(\mathbf{s})$ to be weakly stationary. Then separately define what it means for $Z(\mathbf{s})$ to be isotropic. **[3 MARKS]**

   (b) Give an example of a data set that is likely to be anisotropic. Justify your answer. **[2 MARKS]**

(ii) Consider a geostatistical process with random variables $Z(\mathbf{s}_1)$ and $Z(\mathbf{s}_2)$, where the spatial locations are $\mathbf{s}_1 = (1, 0)$ and $\mathbf{s}_2 = (2, 0)$. Now consider a prediction location $\mathbf{s}_0 = (3, 0)$.

   (a) Letting d denote the distance between 2 points, define the exponential covariance function at distance d with range parameter ϕ, nugget τ^2 and partial sill σ^2. **[2 MARKS]**

CONTINUED OVERLEAF/

(b) Now suppose that $\phi = 1$, $\tau^2 = 0$ and $\sigma^2 = 1$. Calculate the covariance matrix $\mathbf{\Sigma}(\phi, \tau^2, \sigma^2)_{2\times 2}$ for the random variables $(Z(\mathbf{s}_1), Z(\mathbf{s}_2))$ resulting from the exponential covariance function defined above. **[2 MARKS]**

(c) Assuming that $Z(\mathbf{s})$ is a stationary zero-mean process with realisations $z(\mathbf{s}_1) = -0.5$ and $Z(\mathbf{s}_2) = 0.5$, compute the ordinary Kriging predictor for $Z(\mathbf{s}_0)$. **[6 MARKS]**

Hint - It may help you to remember that

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = \frac{1}{ad - bc}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}.$$

(d) Calculate the variance of your ordinary Kriging prediction at $Z(\mathbf{s}_0)$, and hence compute a 95% prediction interval for $Z(\mathbf{s}_0)$. **[3 MARKS]**

(e) Given the location of the prediction point $\mathbf{s}_0$ in relation to the two data points $(\mathbf{s}_1, \mathbf{s}_2)$, is it sensible to make a prediction at $\mathbf{s}_0$ using data $(z(\mathbf{s}_1), z(\mathbf{s}_2))$? Justify your answer. **[2 MARKS]**

2. (i) Consider an areal unit process $\mathbf{Z} = (Z(\mathbf{s}_1), \ldots, Z(\mathbf{s}_n))$, with corresponding neighbourhood matrix $\mathbf{W}$.

(a) Define the k-nearest neighbours method for specifying a neighbourhood matrix $\mathbf{W}$, and give two drawbacks of using this approach. **[3 MARKS]**

(b) Give an example of a situation where the k-nearest neighbours method would be preferable to the commonly used 'sharing a common border' method. Justify your answer. **[2 MARKS]**

(c) Define the Local Indicator of Spatial Association (LISA) based on Moran's I statistic for region i. What can it tell you about the data in region i? **[3 MARKS]**

(d) The LISA was computed for three areas and gave values of: $I_1 = -0.5$, $I_2 = 0.03$, $I_3 = 0.45$. Describe what these three values tell you about the spatial dependence at areas $(1, 2, 3)$. **[3 MARKS]**

(ii) Consider the following areal unit process and model: $\mathbf{Z} = (Z(\mathbf{s}_1), \ldots, Z(\mathbf{s}_n)) \sim \mathrm{N}(0, \mathbf{Q}^{-1})$, where $\mathbf{Q}$ is the precision matrix and hence $\mathbf{Q}^{-1} = \mathbf{\Sigma}$ is the covariance matrix.

(a) What does it mean for the dependence between $(Z(\mathbf{s}_i), Z(\mathbf{s}_j))$ if: (i) the ijth element of $\mathbf{Q}$ equals zero; (ii) the ijth element of $\mathbf{\Sigma}$ equals zero? **[3 MARKS]**

(b) Suppose now that $\mathbf{Z}$ comes from the *proper conditional autoregressive model*. Define $\mathbf{Q}$ for this model, making sure that you define each of the elements in the formula. **[3 MARKS]**

CONTINUED OVERLEAF/

(c) Is the *proper conditional autoregressive model* weakly stationary? Justify your answer. **[3 MARKS]**

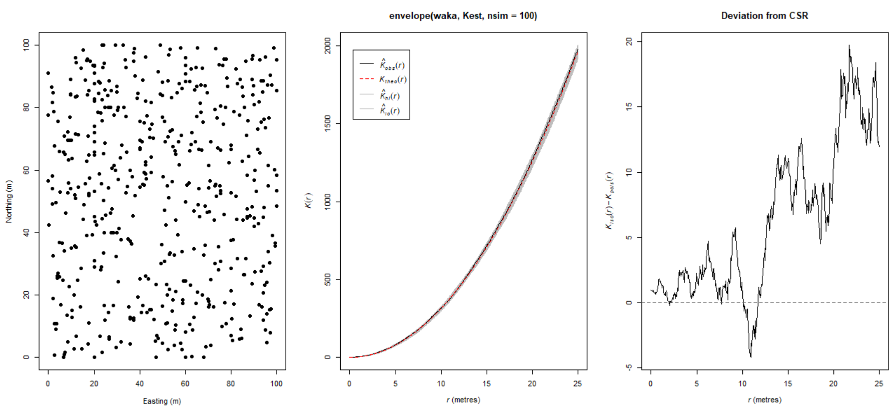
3. (i) Consider a spatial point process $Z = \{Z(A) : A \subset D\}$, where D is the domain of interest.

(a) One hypothesis test for quantifying whether an observed spatial point pattern is completely spatially random is based on quadrat counts. Write down the null and alternative hypotheses for this test, the test statistic, and the distribution of the test statistic under the null hypothesis. **[4 MARKS]**

(b) Consider an observed spatial point pattern with $n = 100$ points across a rectangular domain D. The rectangular domain is then split into 6 quadrats defined by two rows and three columns. The number of points in each of the six quadrats are: 20, 15, 10, 30, 12, 13. Conduct a test of complete spatial randomness and evaluate what it tells you about whether the observed point pattern is completely spatially random. **[4 MARKS]**

(c) Give two downsides of the hypothesis test based on quadrat counts that you have just conducted. **[2 MARKS]**

(ii) The figure below displays the locations of trees in a 100 metre square of the Waka national park in Gabon (left), as well as an estimate of Ripley's K function (middle) , and its deviation from complete spatial randomness (right).



(a) Define Ripley's K function and describe how it can be used to assess if an observed point pattern is completely spatially random. **[2 MARKS]**

(b) From the figure above, do the trees appear to be completely spatially random? Justify your answer. **[3 MARKS]**

CONTINUED OVERLEAF/

(c) You decide to fit an inhomogeneous Poisson process model to the Waka tree data to estimate the extent to which the first order intensity function $\lambda(\mathbf{s})$ varies spatially. Would a parametric or non-parametric model be best for estimating $\lambda(\mathbf{s})$ here? Justify your answer. If you think a parametric model is best, write down a sensible model given the figure above. **[3 MARKS]**

(d) Give two reasons why an ecologist would be interested in estimating $\lambda(\mathbf{s})$ for in Waka forest. **[2 MARKS]**

Total: 60

END OF QUESTION PAPER.