Spatial Data Analysis, Spring term 2021, TAKE-HOME Assignment (due date: 16.05)

INSTRUCTIONS: You may use any books, references, notes and computers. but are not allowed to discuss this home-task with any person other than the instructor until the due date. Be sure to show **all** of your work **in detail**. Please put your *name* and *student no.* in your writeup and SUBMIT all your work, output with the computer codes to the moodle.

1. (Problem given in lab 6)

The data 'swamp.csv' includes locations of 13 different tree species over a (200m $\times$ 50m) rectangular region in a swamp hardwood forest : water tupelo (Nyssa aquatica), bald cypress (Taxodium distichum), etc

Read in the dataset and extract the data values for 'live cypress tree locations' as below.

> swamp <- read.csv('swamp.csv', as.is =T) # data linked with 'swamp.csv'

> your.object <- subset(swamp, live == 1 & sp =='TD') # keep live cypress (Taxodium distichum) locations

a) Create a 'ppp' object for cypress tree patterns using the boundary window: [0,50] x [0,200]. Plot the data to visually examine the distributions for patterns of each data.

> as.ppp(your.object[,c('x','y')], W = c(0,50, 0, 200))

b) Use a kernel method to estimate the intensity of the 'cypress tree' point process. Plot the estimated intensity with an appropriately chosen bandwidth.

c) Test the CSR for the cypress tree data using the K-function. Obtain the simulation envelope to do the significance test.

d) Fit a HPP model for the cypress point pattern. Give the plots of predicted values and standard errors from the model.

e) Fit an inhomogeneous Poisson model where the log intensity = $\beta_0 + \beta_1 x + \beta_2 y$. Give the plots of predicted values and standard errors from the model. Compare this model to the one fit in d).

f) Extract the data for live water tupelo trees : subset(swamp, live ==1 & sp =='FX')

Calculate the cross-K (K_{12}) function and the bivariate $\hat{L}$ to test independence of cypress and water tupelo. Obtain the simulation envelope using the toroidal shift methods. What is your conclusion regarding the relationship between the distributions of cypress and tupelo?

---

2. The data file 'iowa.Rdata' contains 'iowa', SpatialPolygonsDataFrame with information about the number of confirmed COVID-19 cases in Iowa by county observed on 15 March, 22 March, and 13 April 2020. The data frame includes the $\log(N+1)$ transformed variables:

log10pop: log transformed population for each county

logmar15, logmar22, logapr13: $\log(N+1)$ transformed case counts on the dates above

a) Produce the Iowa map showing the log counts on 13 April for each county.

b) Use the shared boundary to define the neighbors of each polygon. Make the plot of neighbor system for each county.

c) Using binary weights, test whether the log transformed counts are spatially correlated. Repeat with the row standardized weights.

d) We want to see whether there is an association between the Mar 22 count and Apr 13 count (both log transformed), after accounting for spatial correlation with neighbors.

We will use 4 different (spatial) linear models to explore such an association: independence, SAR binary weights, SAR row-standarized weights, CAR

Compare each model fit based on the AIC criteria and tests of spatial autocorrelation on the residuals.

e) For each model, redo the spatial regression analysis by adding the variable 'log10pop' to the previous models and check if it improves the model estimation.

f) Make the grayscale plots of the fitted values and residuals from the best model chosen above. Draw a brief conclusion regarding the existence of any association, etc.

---

3. Consider a two-dimensional, intrinsically stationary random field, having a spherical semivariogram with range 0.5 units. Suppose this random field is observed at data locations which form a square grid spacing 1.0 units.

a) What would the sample semivariogram of the data tend to look like? (Draw a picture.)

b) Note that measurement error can cause a nugget effect in the sample variogram. What does this example suggest about another possible cause of a nugget effect in the sample variogram?

---

4. Consider a second-order stationary 3-dim random field,

$$Z(x, y, z) = \mu + \epsilon_{ME}(x, y, z) + \epsilon_{NE}(x, y, z)$$

where $\epsilon_{ME}(x, y, z) = \delta(x, y, z) + \eta(x, y)$; $\delta(x, y, z)$ is a measurement error process with covariance function

$$C_\delta(\boldsymbol{h}) = \begin{cases} \sigma_\delta^2 & \boldsymbol{h} = \boldsymbol{0} \\ 0 & \text{otherwise} \end{cases} \qquad C_\eta(\boldsymbol{h}) = \begin{cases} \sigma_\eta^2 & \boldsymbol{h} = \boldsymbol{0} \\ 0 & \text{otherwise} \end{cases}$$

$\epsilon_{NE}(x, y, z)$ is an isotropic process representing the residual variation not accounted for by measurement error. Assume that all three components of the residual process are uncorrelated.

Compute the semivariogram of $Z(\cdot)$ in the "z-direction" ($\gamma(0, 0, h)$) and in all other directions. What form of anisotropy does this random field exhibit?