Spatial Data Analysis, Final EXAM (May 7, 2021)

INSTRUCTIONS: Show your work on all questions. Incorrect answers with sufficient work will get (enough) partial credit, and correct answers with insufficient work may not get full credit. This exam includes 4 main problems with 3 small (alphabet-ordered) extra-credit problems.

1. a) Consider a random process $\{Y(\boldsymbol{s})\}$ with $\text{Cov}[Y(\boldsymbol{s}), Y(\boldsymbol{t})] = \sigma^2(\|\boldsymbol{s}\| + \|\boldsymbol{t}\| - \|\boldsymbol{s} - \boldsymbol{t}\|)$, and $\text{E}[Y(\boldsymbol{s})] = 0$.

   Is $\{Y(\boldsymbol{s})\}$ intrinsically stationary? Is it second-order stationary? Justify your answer.

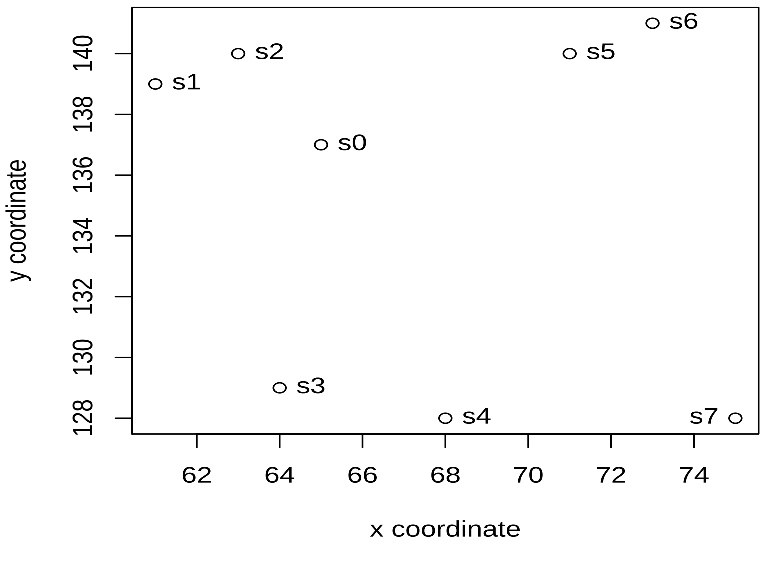
   b) Why does one need to choose a semivariogram model to fit to the empirical semivariogram estimates not using empirical semivariogram estimates directly?

   c) In estimation of the semivariogram model parameters (sill, range, nugget), clearly the ordinary least squares (OLS) estimation is not appropriate to use. Explain briefly why it is not appropriate.

2. Find the proximity (spatial weight) matrix W using the rook's definition for the data:

Z_1	Z_2	Z_3
Z_4	Z_5	Z_6

3. Consider the following data given on the graph below. $Z(\boldsymbol{s_1}), Z(\boldsymbol{s_2}), ..., Z(\boldsymbol{s_7})$ are the observed values of the Z process which is described by the exponential semivariogram $\gamma(h) = c_0 + c_1(1 - e^{-h/\alpha})$, with $c_0 = 0, c_1 = 10, \alpha = 3.33$.



In R, > x=c(61, 63,64, 68,71,73,75)

> y= c(139,140, 129,128,140,141,128)

> z= c(477,696,227,646,606,791,783)

a) Construct the variance-covariance matrix of the vector $\boldsymbol{Z} = (Z(\boldsymbol{s_1}), Z(\boldsymbol{s_2}), ..., Z(\boldsymbol{s_7}))^T$. Recall that $\gamma(\boldsymbol{h}) = C(\boldsymbol{0}) - C(\boldsymbol{h})$.

b) Compute the variance of $\tilde{Z}(\boldsymbol{s_0}) = 0.2Z(\boldsymbol{s_1}) + 0.2Z(\boldsymbol{s_2}) + 0.2Z(\boldsymbol{s_3}) + 0.1Z(\boldsymbol{s_4}) + 0.1Z(\boldsymbol{s_5}) + 0.1Z(\boldsymbol{s_6}) + 0.1Z(\boldsymbol{s_7})$ using R.

c) One goal is to predict the value $Z(s_0)$ at location $s_0 = (65, 137)$. Using the variogram, compute the matrix Γ_o and the vector $\boldsymbol{\gamma_o}$ needed for the ordinary kriging system of equations. (Or compute the matrix C and the vector c based on the covariance function where $Cw = c$).

Obtain the ordinary kriging weights, corresponding predicted value and the prediction variance.

d) Repeat the kriging prediction conducted in c), but with the nugget of 10. How do the kriging weights, corresponding predictions and prediction variance change? Discuss briefly.

e) Universal kriging: Assume a linear trend as a function of the coordinates x an y. Obtain the universal kriging weights, corresponding predicted value and the prediction variance. Compare the results with those in c).

4. a) What are the measures (functions) to characterize the global feature and spatial interaction (respectively) for the point pattern data? Give one appropriate (commonly used) tool (estimator) for each measure.

   b) For the rotavirus data spread in Germany (stored in the moodle), perform the test of CSR (complete spatial randomness) using the quadrat test, G function and Ripley's K function with simulation envelope. Make brief comments on the test results.

   c) Fit three candidate models (of your own choice) for HPP, IPP, and PCP to fit the data. Compare these fitted models and choose the best one among them. Provide your reasoning briefly.