Exercise Day 2

Exercise 1

Load the river blindness data, `LiberiaRemoData.csv`. Consider the empirical logit transformation of prevalence, given by

$$\tilde{Y}_i = \frac{Y_i + 0.5}{n_i - Y_i + 0.5},$$

where Y_i and n_i are the number of positive cases and number of tested individuals at location x_i.

1. Fit a linear geostatistical model to the empirical logit, i.e.

   $$\tilde{Y}_i = \alpha + \beta \log\{d(x_i)\} + S(x_i) + Z_i$$

   where $S(x_i)$ is a stationary Gaussian process with exponential correlation function and Z_i are i.i.d. Gaussian variables.

2. Use the model from the previous question to predict nodule prevalence across Liberia and display the exceedance probability for a 20% threshold.

3. Compare the results of the predictions with those of the Binomial geostatistical model presented in class. What do you conclude?

Exercise 2

Load the data `anopheles.csv` containing number of Anopheles mosquitoes in Southern Cameroon.

1. Fit a non-spatial Binomial model for the number of An. gambiae (`An.gambiae`) using `elevation`, as an explanatory variable. Finally, use this model to predict the number of mosquitoes using the prediction grid and elevation values in the file `anopheles_elevation.csv`.

2. Carry out the test for residual spatial correlation based on the variogram. Is there evidence of unexplained spatial variation?

3. Fit a Poisson geostatistical model to the counts of An. gambiae and carry out spatial prediction as in 1.

4. Compare the predictions from 1 and 3. Is there any difference? What would be the rationale for using a geostatistical model in this case?