ASSIGNMENT 2
PM569 Spatial Statistics, Fall 2019

Due October 8, 2019

Fitting semivariograms/covariance functions with simulated data

1. Simulate a Gaussian random field using the `grf()` function in the `geoR` library with a set of self-chosen spatial parameters for the matern function $(\kappa, \tau^2, \phi, \sigma^2)$. Consider this as your "true" spatial process. Recall that in R, σ^2 is the sill minus the nugget.

    a. Plot the cloud and binned empirical variogram and provide an eyeballed estimate of the spatial parameters. Do they conform with your intuition given the spatial process you simulated?

    b. Using the Matern and Gaussian function, fit the semivariogram for these simulated data estimating the parameters using OLS, WLS, ML, and REML. Add each fitted line to the empirical semivariogram. Put your results in a table, providing the parameter names and their true values, each of the estimated values, and a fit measure (SSE, AIC or BIC). Provide a brief description of each of the 4 methods, and discuss well they each fit your simulated spatial process.

    c. With the same parameters, simulate a random field 2 more times and save the output dataset each time. Using your preferred method from b., estimate the parameters again. Do the parameter estimates from these models depend greatly on the simulated dataset?

Fitting semivariograms/covariance functions with real data

For this question we will use **wind speed** from the Hurricane Dorian dataset (HW1). Please use projected coordinates for this question.

2. Using WLS and maximum likelihood (ML or REML), fit the exponential, Gaussian, and Matern functions. Create a table with all spatial parameter estimates from both methods and also record the sums of square error or AIC/BIC. Describe the fitted parameters, and compare the models. Choose what you feel is the best model for the data.

Kriging

3. With your chosen model from 2., perform ordinary kriging on a grid. Plot (separately) the kriged estimates and standard errors. Please report all parameter estimates and describe in words what this kriging method is doing.

4. Determine whether a linear or quadratic trend fit is more appropriate. Then, with your chosen model from 2., perform universal kriging (with trend) on a grid. Create a map using leaflet of the kriged estimates and standard errors. Please report all parameter estimates and describe in words what this kriging method is doing.

Cross Validation

5. Split the Dorian data into a test and a training set (70-30%) and re-do 4. on the training set, and predict with the test set. Assess kriging performance by comparing observed and predicted wind speed via the cross validation MSE and R^2 .

Spatial Regression and Interpolation

6. Fit three different spatial models for wind speed: 1) `lm()`, 2) `gls()`, and 3) `gam()`. For `gls()`, choose the best covariance/correlation structure between `corRatio`, `corGaus` and `corSpher` that includes an interaction term in the trend component. For `gam()`, use the default number of knots for the thin plate spline. For all models please examine the other variables in the dataset at possible predictors.
   Describe the three modeling approaches and explain their differences. Where possible, compare the β coefficients of x and y and their standard errors as well as the model fits using `AIC()` on each model object.

7. Run `gam()` again using a larger number of knots (try 75% of the data size). Compare with your `gam()` results using the default settings. What is happening by increasing the number of knots? For spatial-only gam models, use `predict.gam()` and map the predictions from `gam()` models with the default number of knots and increased number of knots (and their standard errors). Describe the differences in your maps. Why can't we create a map using gam models with covariates?

8. Plot the binned semivariogram of the residuals from the `gam()` models, the `lm()` model, and your preferred `gls()` model. Examine and describe any residual spatial pattern from these models.

9. Interpolate wind speed using inverse distance weighting (IDW) using two different values of ρ and make a map of your results. Explain what changing ρ does to your estimates. Discuss any advantages or disadvantages of the IDW method.