Homework 1
STA465/STA2016/ENV1112 Spatial Data Analysis, Fall 2024

Due Sunday, September 29, 2024 (end of day)

A. Everyone: Spatial Data and Spatial Computations (30 points)

The goal of this part of the assignment is to get familiar working with different types of spatial data, projections and spatial computations. The zip file hw1-data.zip contains Three datasets: Toronto_Neighbourhoods.shp, TTC_SUBWAY_LINES_WGS84.shp, and toronto-crimes-06-2024.csv.

1. (5 pts) Read in all files, summarize what type of spatial data each represents, and create a single leaflet map showing all of the data. Color the subway lines by route, add labels so when you interact with the map the subway route and neighborhood name appear. Use `addLayersControl()` to be able to click on and off layers on your map.
2. (1 pt) Do each of your datasets have a datum? Briefly explain what this is and if needed apply a single datum type to all 3 datasets.
3. (1 pt) Look up the appropriate UTM zone for Toronto and apply this projection to the data (hint: use epsg.io to find the right code). Briefly explain why this projection step is necessary.
4. (3 pts) Calculate the area of each Toronto neighborhood and summarize. Is there a correlation between area and population?
5. (8 pts) Using a spatial join, determine which Toronto neighborhoods had the most crimes in June 2024. Create two maps of your join: a) showing the counts of crimes in each neighborhood and b) showing the per capita crime rate (count/population). Make a legend for each map and interpret any spatial patterns you might observe.
6. (6 pts) Create a 1 km buffer around the subway lines. Visualize the 1 km buffer on a map to check if it's correct. By doing another spatial join, what proportion of the crimes occurred within 1 km of a subway line?
7. (6 pts) Which neighborhoods intersect the subway lines? Do the intersecting neighborhoods have more or fewer crimes than those that do not have a subway?

B. STA Students Only: Theoretical Semivariograms and Covariance (20 points)

1. (8 pts) For the following theoretical semivariogram functions:

   – Linear: $\gamma(h) = \tau^2 + \sigma^2 h$ if $h > 0; 0$ otherwise

   – Power model: $\gamma(h) = \tau^2 + \sigma^2 h^\lambda$ if $h > 0; 0$ otherwise, $\lambda > 0$

   – Rational quadratic: $\gamma(h) = \tau^2 + \sigma^2 \frac{h^2}{(1+\phi h^2)}$ if $h > 0; 0$ otherwise

   – Powered exponential: $\gamma(h) = \tau^2 + \sigma^2 (1 - exp(-|\phi h|^\lambda)$ if $h > 0; 0$ otherwise, $\lambda > 0$

   – Wave: $\gamma(h) = \tau^2 + \sigma^2 (1 - \frac{sin(\phi h)}{\phi h})$ if $h > 0; 0$ otherwise

   a. Plot the semivariograms (on the same plot) fixing $\tau^2 = 0.5, \sigma^2 = 4, \phi = 6, \lambda = 0.5$ but choosing your own distances.

   b. Using the plot from a), provide a description of the differences in the four models despite having the same fixed parameter values. Also explain whether or not each semivariogram corresponds to a stationary process.

   c. For the two power models, vary the value λ and plot the different semivariograms. What does the power parameter λ represent?

   d. For the rational quadratic and wave models, vary the values of τ^2, σ^2, ϕ (one at a time) and describe how the semivariogram changes.

2. (6 points) Write out and plot the **covariogram** (covariance as a function of distance) for the exponential, spherical and Gaussian functions shown in class.

3. (3 points) The Matern covariance function is widely used in the spatial statistical literature. The function is:

$$C(h) = \sigma^2 \frac{1}{\Gamma(\kappa) 2^{\kappa - 1}} (\sqrt{2\kappa} \frac{h}{\phi})^\kappa K_\kappa (\sqrt{2\kappa} \frac{h}{\phi})$$

   Where κ is a smoothness parameter, Γ is a gamma function, and K_κ is a modified Bessel function. Using the functions gamma() and besselK(), write the above Matern covariance function in R and test and plot it with your choice of parameter values. What do you see changing for small κ (i.e. $\kappa \to 0$) versus large κ (i.e. $\kappa \to \infty$)?

4. (3 points) Using the dorian.csv dataset provided (see details below in part C.), import the data, project the coordinates to UTM, and plot the binned empirical semivariograms for windspeed and atmospheric pressure.

C. ENV Students Only: Empirical Semivariograms (20 points)

For this question we will use meteorological data from weather stations at the time of Hurricane Dorian (North Carolina), September 6, 2019 (averaged over hourly measurements 10am-3pm). We will focus on wind speed (m/s) and atmospheric pressure (millibars) to examine the strength and location of the hurricane. The data were acquired from the National Oceanic and Atmospheric Administration's National Weather Service ftp (ftp.ncdc.noaa.gov/pub/data/noaa/).

1. (6 points) Perform exploratory data analysis: examine the data distributions of wind speed and atmospheric pressure and provide summary statistics. Create two maps in leaflet showing the locations and color gradients for the values of the meteorological parameters. Please describe any spatial trends that can be visualized in the data.

2. (0 points) Project the latitude and longitude to UTM coordinates. Use these x,y values in the subsequent questions.

3. (4 pts) Plot empirical semivariograms (robust) for wind speed and atmospheric pressure. Try two different semivariograms for each variable (e.g. change max distance, number of bins).

4. (4 pts) Provide an eyeball estimate of the nugget, sill and range and plot a curve of a theoretical semivariogram (your choice of which function to use) on top of your preferred binned semivariogram.

5. (2 pts) Discuss your findings by comparing the spatial aspects of each variable. Do you think one or the other is a better indicator of the hurricane's strength and scope?

6. (2 pts) Determine if there is a linear spatial trend in wind speed or pressure.

7. (2 pts) Determine if there is anisotropy in either variable.