Agricultural Sustainability in the Anthropocene: Weekly Journal

Leila Uy 100401717

Meeting: May 4th 2021

For this research opportunity, I want to cultivate my research skills, improve my academic writing, and expand my knowledge of GIS and programming. My main focus throughout the summer session is to use remote sensing and parallel programming to make a quick and efficient geographical Spatio-temporal clustering algorithm for large datasets.

I had a brief encounter with clustering in GGR376 (Spatial Data Science II). Coded in R, we used IDW interpolation, kriging, and skater clustering to find clusters in air pollutants. I used this knowledge to understand in-depth the article by Kumar et al. They used K-means clustering, which uses iterations to navigate the best centroid for the eco-regions. This process takes a lot of computation and a majority consist of Euclidean distance calculations. They reduced the algorithm time by getting rid of unnecessary evaluations using triangle inequality.

Additionally, the article by Ellis et al. raises the problem that existing systems for representing global patterns ignore or oversimplify human influence. A solution they proposed is an anthropogenic biome map that used population density as the main variable to indicate human interactions with ecosystems. They identified natural groupings through cluster analysis based on populations, land-use and land-cover characteristics, and regional distribution.

I am interested in implementing a parallel K-means clustering algorithm that accounts for human influence. While reading the articles a few questions came into mind:

- Should we use population density as a variable in our k-means clustering?

- Are there more calculations we could optimize in the algorithm?

- What is the upper boundary or limit of the algorithm?

- Is there a limitation to purely data-based analysis? If so, how do we find the balance between data and expert opinion?

- What are the best ways to do accuracy assessment? Ground collection? Satellite imagery?

Olson et al. proposed that in the past, maps of global biodiversity were ineffective planning tools because they divided the earth into coarse biodiversity units. The large units made it difficult to discern smaller but highly distinctive areas that were useful for conservationists. Through parallel programming and satellite imagery, I am hoping we can resolve this problem.