POPULATION MODELING IN MEXICO WITH WIRELESS DEVICE LOCATION AND NIGHTTIME LIGHTS DATA

Nanshan Li, Bob Chen, Alex de Sherbinin, Kytt MacManus, Greg Yetman
CIESIN, The Earth Institute at Columbia University

Motivation

- Find correlation between data and population to establish a baseline for changes in population
- Applications: response to natural disasters, conflicts

NASA Visible Infrared Imaging Radiometer Suite (VIIRS)

Image 1: Baja Norte (left) and Federal District (right), the two states in Mexico for which the luminosity values were processed and analyzed

- 608 processed satellite images of the nighttime lights, taken from the equator at approximately 1am nightly from Jan 2017 - Sep 2018
- Images are processed with algorithms developed to correct for atmospheric, BRDF and seasonal effects (Román et al, 2018).

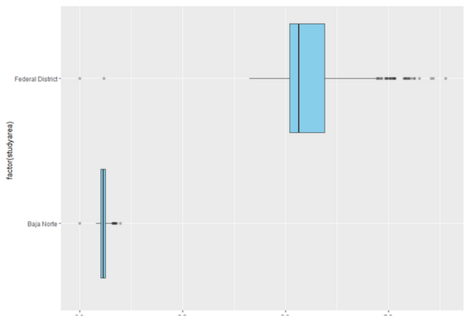

Image 2: Mean and variance in Federal District (Mexico City) is significantly greater than in Baja Norte, its less populous counterpart

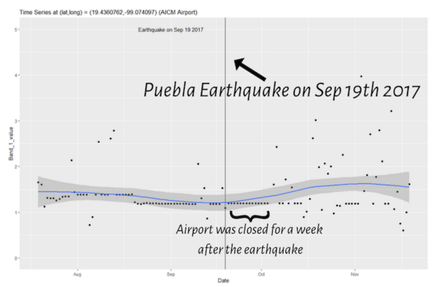

Image 3: Time series of the luminosity values at the airport around the earthquake

- Values in each raster cell is frequently imputed by the algorithm, cannot be analyzed on a daily scale
- Correlation of brightness of lights with population is suspect

Device Location Datasets (Location pinpointing with Wi-Fi, GPS and Cell Signals)

- Includes location stamps from devices across multiple platforms and providers, including Apple, Linux, Windows, Android and Blackberry
- First set of data: MAC & SSID counts aggregated by 30m x 30m tiles in Mexico
- Second set of data: MAC ID & Location request counts aggregated by hour and 100m x 100m tiles, from Sep 30 - Oct 8 2018
- Analysis done on a subset of the data from Federal District

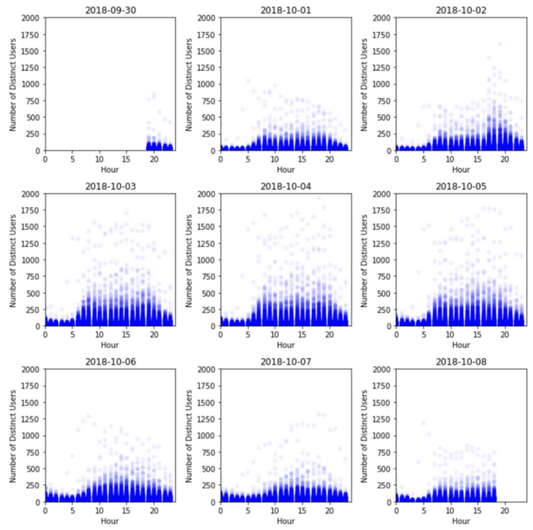

Image 4 : Scatterplot of entire time series for the time period (each dot: 1 admin block)

Device usage pattern on weekends is distinct from the device usage pattern on weekdays: 3 peaks on weekdays vs 2 on weekends.

* Anomalous jump in usage on the evening of Oct 2, 2018 can be attributed to the rally held in remembrance of the 50th anniversary of the Tlateloco Massacre

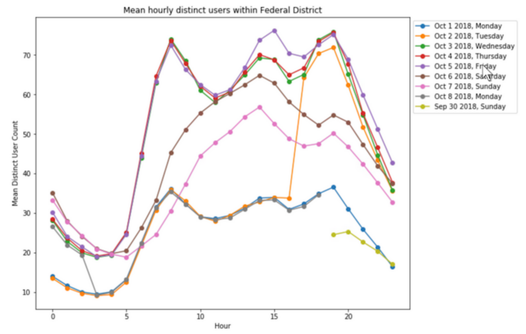

Image 5 (right): mean hourly distinct users within the Federal District

k-means clustering with k = 2 after scaling and PCA on geoprocessed dataset 2 to separate census block tracts based on usage patterns (silhouette score of 0.146)

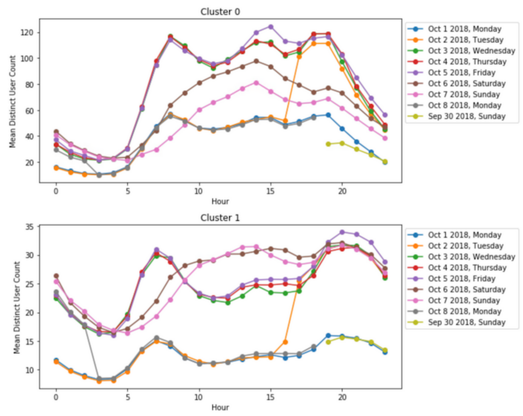

Image 6: Hourly mean distinct users for clusters 0 and 1: Cluster 0 may represent business districts, with commuting times and lunch-time activity representing the three peaks, while areas with higher relative weekend activity (cluster 1) may represent residential districts. Limited validation using neighborhood descriptors from Wikipedia suggests this may be the case.

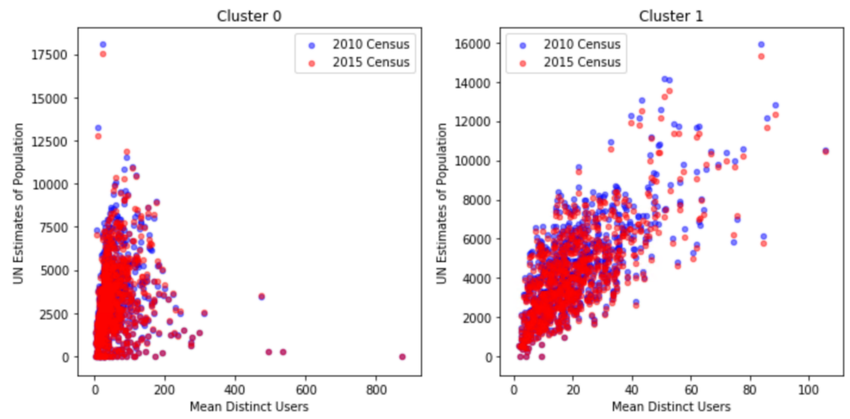

Image 7: Scatterplot of mean distinct users in each cluster with UNE of population from 2010 census; cluster 1 correlates better with UNE population for both 2010 and 2015

Conclusion and future steps

- Device location data shows greater promise than VIIRS dataset for modeling population on a smaller spatial and temporal scale
- Clustering was a viable approach to separate census tract blocks of different use-type; will experiment with more clusters to account for mixed-use neighborhoods
- Next step is to derive correlation with both ambient (daytime) and resident (nighttime) population.
- Also looking into correlation with other demographic attributes to see if they can potentially be used as clustering features or labels.

1 - Wang, Z., Roman, M. O., Sun, Q., Kalb, V., Macmanus, K., Ryan, R. E., . . . Helder, D. (2018). Nasas Black Marble Product Suite: Validation Strategy. IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium. doi:10.1109/igarss.2018.8518106