Analyzing Houston Flooding Using Unsupervised Machine Learning

RICE

NAKATANI FOUNDATION for advancement of measuring technologies in biomedical engineering

Etsuko Ishii[1,2], Isaac Dykeman[3], Zhouhan Chen[3], Devika Subramanian[3]

[1]Department of Mathematical Engineering and Information Physics, the University of Tokyo, Japan

[2]Nakatani RIES: Research and International Experiences for Student, Nakatani Foundation, Japan

[3]Department of Computer Science, Rice University, U.S.A.

東京大学 THE UNIVERSITY OF TOKYO

Flooding in Houston

Objective

This research aims at establishing a model which

- Predicts flooding at the home level in Houston based on integrating a range of readily available data sources: rainfall, 311 calls, flood plain designation, digital elevation models
- Uses unsupervised machine learning techniques to identify the key determiners of home flooding.

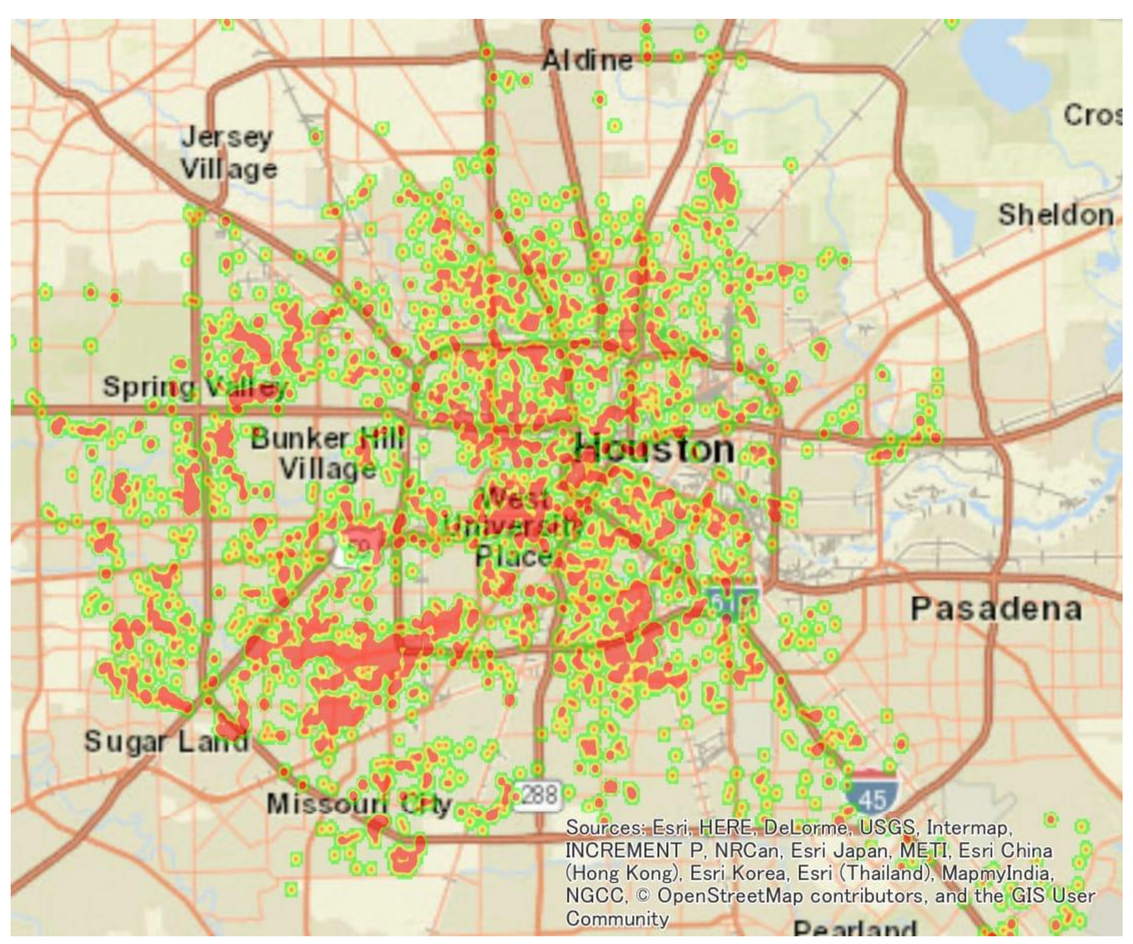


Fig. 1: 311 flooding 311 call density map of Houston in 2011-2015. These are regions which are more likely to flood. Several locations have experienced more than 5 floods in 5 years.

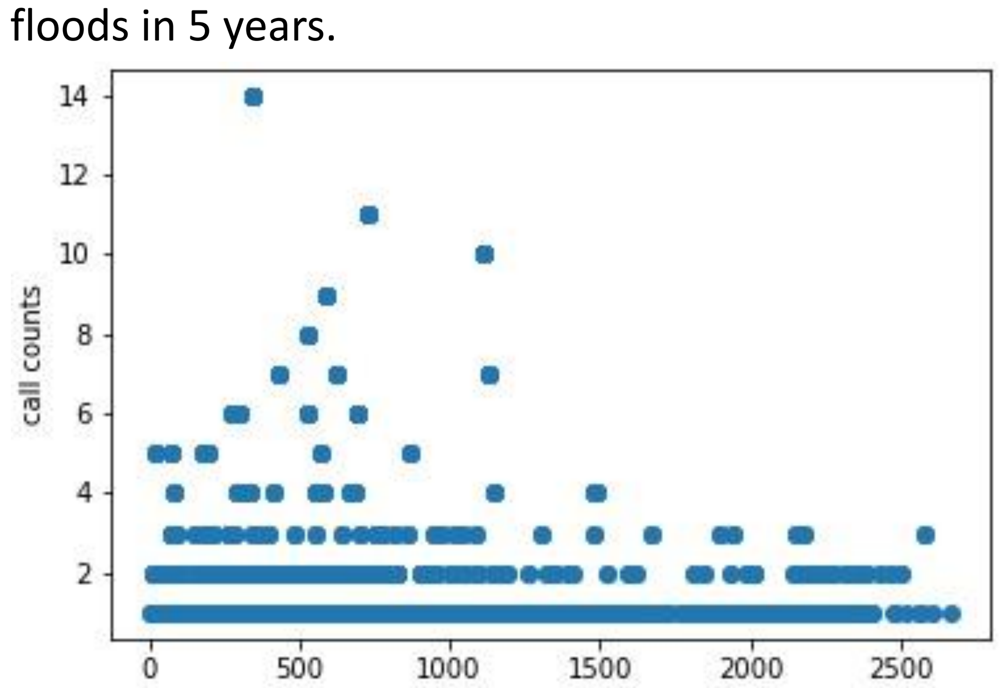
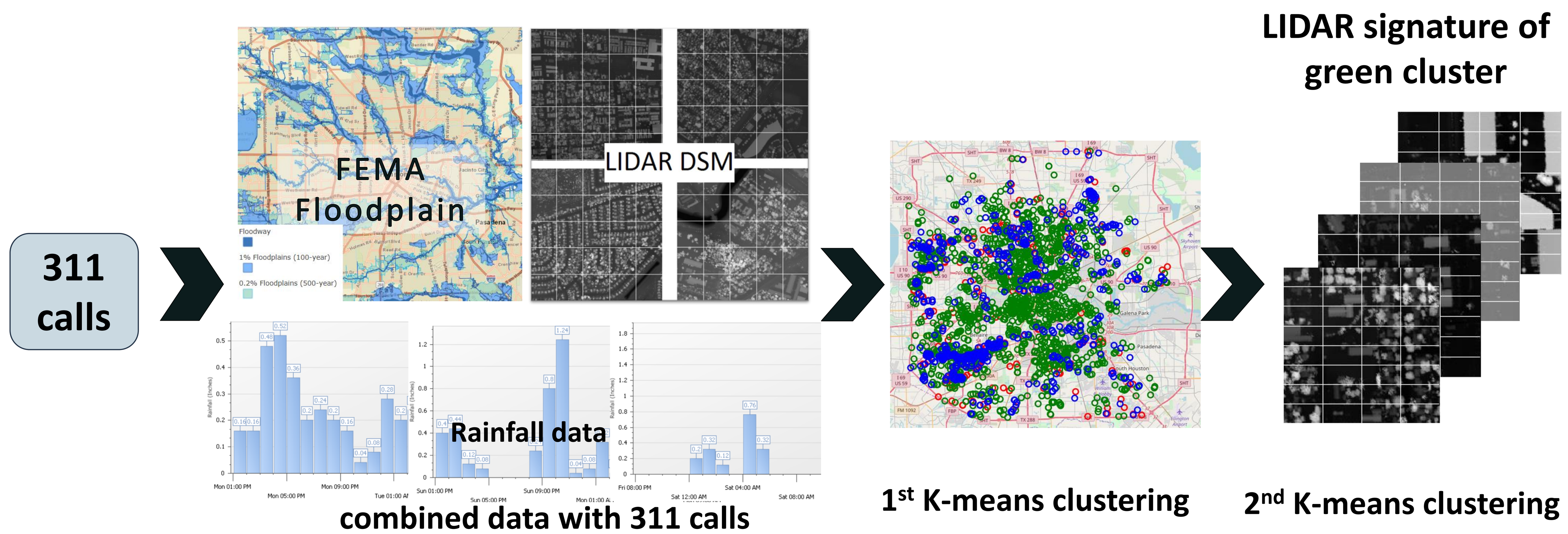


Fig. 2: For homes **not in a designated flood plain**, scatterplot of distance from the nearest bayou and 311 call count in 2013-2015. **Note that 51.2% of homes not in designated flood plains experienced more than one flood in 2013-2015.**

Unsupervised Machine Learning



Data Used (~5000 calls)	
311 calls	flooding issues reported to city of Houston in 2013-2015 with lat/long coordinates
rainfall	total rainfall day of the nearest rain gage to 311 call location installed by Harris County Flood Control District
bayou distance	distance from the nearest bayou to 311 call locations
floodplain	Floodplain designation defined by FEMA for 311 call location
LIDAR DSM	the surface model of the earth that includes all buildings or plants on it around 311 call location

Goal: identify groups of 311 call locations based on rainfall, bayou distance, floodplain, call frequency by K-means clustering

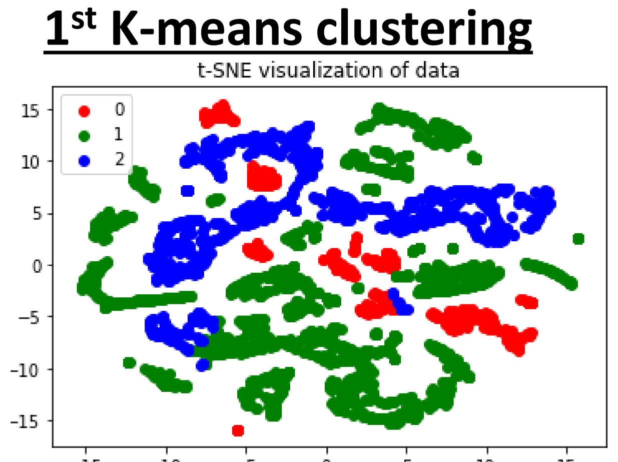
- Normalized input data (5000 x 8 features)
- Classified into three clusters (elbow method)

2nd K-means clustering: identify LIDAR signature of unusual cluster uncovered by 1st K-means clustering

- Extracted LIDAR DSM data for those 311 call locations
- Scaled to lie between given minimum and maximum value
- Classified into four clusters by K-means

tSNE/K-means Clustering Results

1st K-means clustering



Cluster A: very close to bayou, in flood plain, more than 2 floods on average

Cluster B: far from bayou, not in flood plain, more than 2 floods on average

Cluster C: close to bayou, in flood plain, more than 1 flood on average

	bayou distance(m)	call counts	floodplain	percentage
Cluster A	42.6	2.35	0.400	15.0%
Cluster B	**768**	**2.78**	**0.109**	**54.5%**
Cluster C	140	1.58	0.368	30.4%
average	468	1.82	0.232	

Fig. 3: Scatterplot of 311 calls in each cluster color visualized by t-SNE.
Table 1: average feature values of each cluster.

2nd K-means clustering

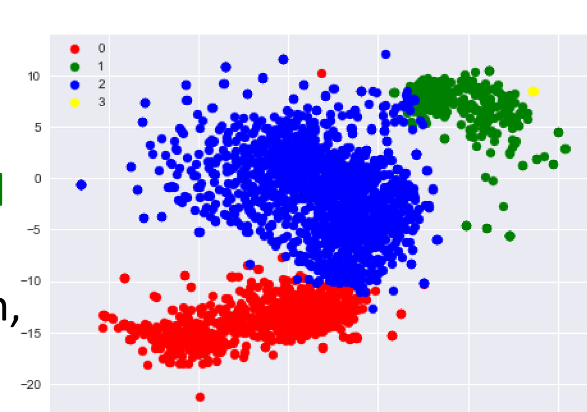
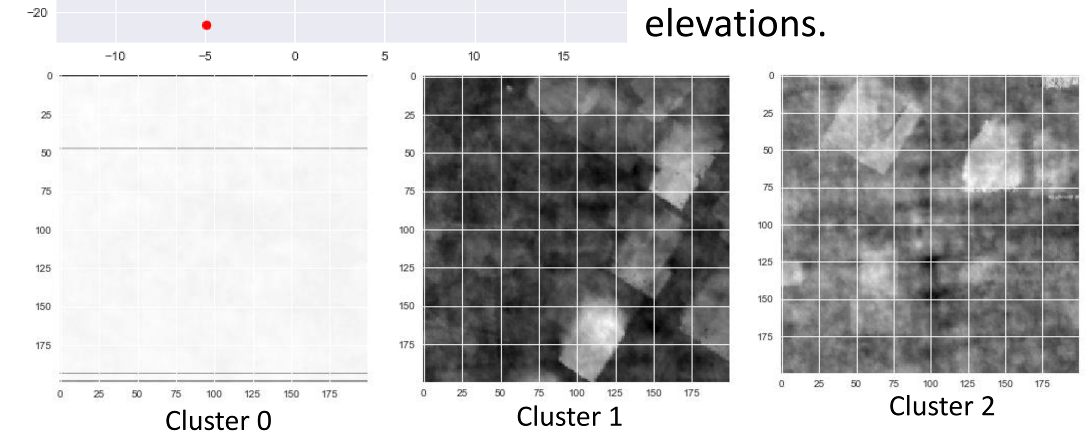


Fig. 4(left): Scatterplot of cluster B visualized by t-SNE.

Fig. 5(below): LIDAR DSM mean images of each cluster in 200m*200m cell. 311 call locations at relatively lower elevations.



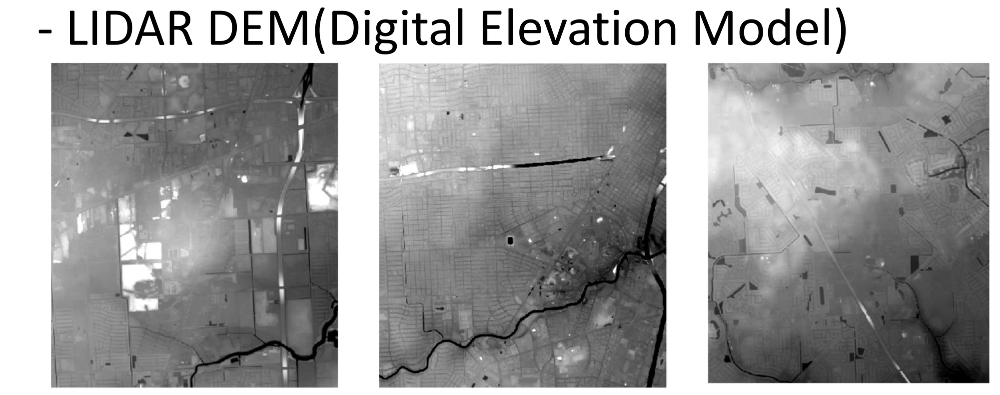
Cluster B was successfully classified into 4 clusters, two of which are shown above.

Conclusions

- Unsupervised machine learning identifies a cluster of over 50% of homes that have experienced multiple floods in Houston during 2011-2015 that are far from a bayou and that are not in FEMA designated flood plains.
- LIDAR analysis of these homes (with a 200mx200m tile around home) reveals that these regions are characterized by high density development and proximity to large tracts of land at low elevation.
- Further analysis of these homes at different scales (50mx50m to 500mx500m) is need to identify key factors that cause flooding.

Future Work

- 2nd K-means clustering by other features
  - LIDAR DEM(Digital Elevation Model)



  - road features
  - water pipe/ drainage/ sewage network data
  - land use features
- Establishing a fast and accurate flood prediction model at home level using diverse data sets rather than by running detailed physical simulations.

References

1. Loftis, J., Wang, H., Hamilton, S., and Forrest, D. (2014). Combination of Lidar Elevation, Bathymetric Data, and Urban Infrastructure in a Sub-Grid Model for Predicting Inundation in New York City during Hurricane Sandy., *eprint arXiv:1412.0966*, 12/2014.
2. Loftis, J., Wang, H., De Young, R., and Ball, W. (2016). Using Lidar Elevation Data to a Topobathymetric Digital Elevation Model for Sub-Grid Induation Modeling at Langley Research Center., *Journal of Coastal Research*, 76(sp1):134-128.
3. Harris County Flood Control District(HCFCD). (2017). Flooding Education Mapping Tool. http://www.harriscountyfemt.org/.
4. The City of Houston. (2017). 311 Flooding Cases. http://www.houstontx.gov/heatmaps/.
5. FEMA Flood Map Service Center. (2017). Floodplain Maps. https://msc.fema.gov/portal/productAvailability/.