Leveraging Machine Learning and Twitter Data to Identify High Hazard Areas during Hurricane Harvey

Antonia Sebastian[1,2,*], William Mobley[1], Wesley Highfield[1], Samuel D. Brody[1] *corresponding author: asebastian@tamu.edu

[1] Center for Texas Beaches & Shores, Department of Marine Sciences, Texas A&M University at Galveston, Galveston, Texas USA
[2] Department of Hydraulic Engineering, Delft University of Technology, Delft, the Netherlands

TEXAS A&M UNIVERSITY GALVESTON CAMPUS

CTBS The Center for Texas Beaches and Shores

GC43D-1374

The timely understanding of flood extent is critical information for emergency managers during disaster response. Search and rescue operations require a boundary to reduce fruitless efforts and to prioritize critical or limited resources. However, high-resolution aerial imagery is often unavailable or lacks the necessary geographic extent, making it difficult to obtain real-time information about where flooding is occurring. Volunteered geographic information (VGI) is a subset of crowdsourcing and can be used to disseminate spatially relevant information or request help. In this study, we present a novel approach to map the extent of urban flooding in Harris County, TX during Hurricane Harvey (August 25-31, 2017) and identify where people were most likely to need immediate emergency assistance based on a subset of crowdsourced SOS requests. Using the machine learning software Maximum Entropy (MaxEnt), we predict the spatial extent of flooding based on several physical and socio-economic characteristics. We compare the results against two alternative flood datasets available after Hurricane Harvey (i.e., Copernicus satellite imagery and fluvial flood depths estimated by FEMA), and we validate the performance of the model using a 15% subset of the rescue requests and Houston 311 flood calls. We find that the model predicts a much larger area of flooding than was shown by either Copernicus or FEMA when compared against the locations of rescue requests, and that it performs well using both a subset of rescue requests (AUC 0.917) and 311 calls (AUC 0.929).

STUDY AREA & BACKGROUND

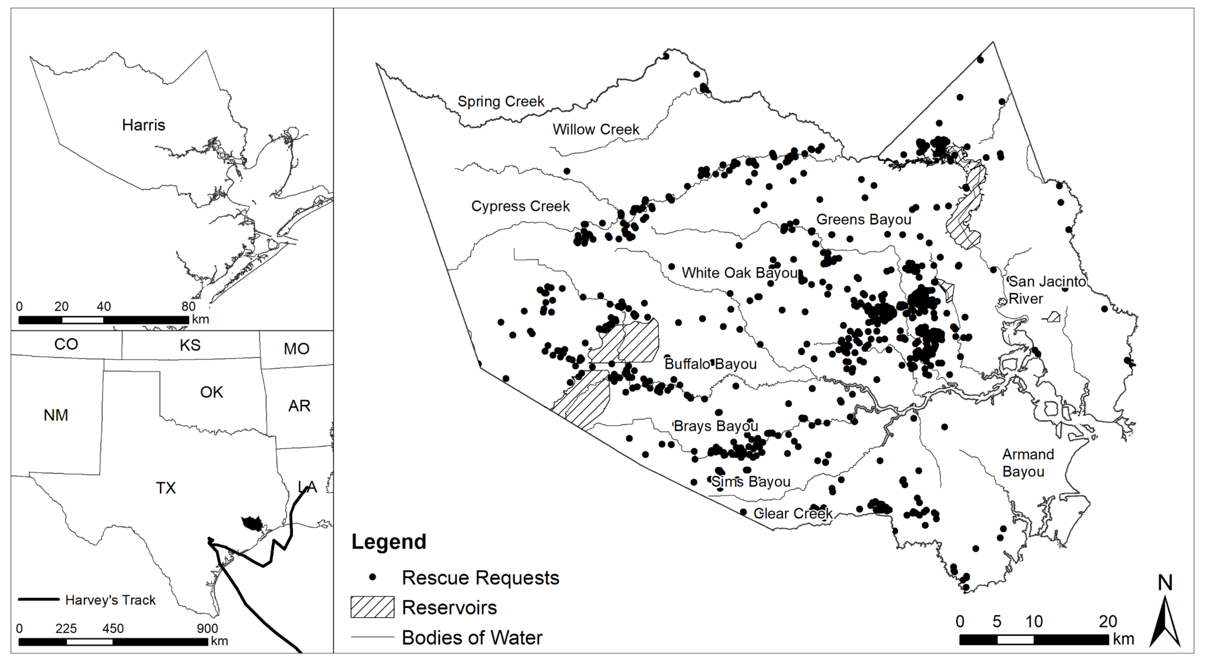


FIGURE 1 Maps showing the location of Harris county relative to track of Hurricane Harvey as the storm approached the Texas coast (left) and rescue requests made in Harris County between August 25 and 30, 2017 (right)

- Hurricane Harvey made landfall as a Category 4 near Rockport, TX on August 25, 2017
- Est. return period between 1/1,000 and 1/9,000 years [1]
- Highest recorded rainfall 1,539 mm (60.58 in) near Nederland, TX [2]
- Est. damage $125 billion USD [2]
- Largest disaster response in state history; est. >100,000 high water rescues [3]
- 800,000 households applied for disaster assistance [3]

OBJECTIVES

1. Utilize the species distribution model (SDM), MaxEnt, to predict probability of flooding at large scales in near-real time based on the locations of Harvey rescue requests
2. Estimate flood extent given a 95% test prediction cutoff
3. Test the performance of the model against a subset of rescue requests, 311 calls, and flooded roadways
4. Compare the model results against Copernicus satellite imagery and post-event riverine flood depths from FEMA
5. Reduce Type II Errors: failure to identify people who are impacted by flooding

DATA & METHODS

MaxEnt is an ecological model widely used to predict the distribution of a species [4-7]

- Fits the distribution that is closest to uniform (i.e., maximizing entropy) given a series of independent variables as constraining features
- Predicts the probability of an event based on point observations of occurrence and background environmental data
- Environmental data (i.e., independent variables) may be continuous, categorical, or binomial

TABLE 1 Independent variables included in the MaxEnt model

	Range	Mean	SD	Source	Data type	Measurement
Topologic variables						
Elevation	0-98.99	25.47	17.76	NED	Continuous	Meters relative to NAVD88
Flow accumulation	0-2,422	5.95	31.88	NED	Continuous	Number of contributing raster cells based on flow direction network
Watershed	—	—	—	HCFCD	Discrete	22 sub watersheds as defined by HCFCD
Distance to coast	0-91,482	33,440	22,279	NHD	Continuous	Meters to nearest coastline
Distance to stream	0-4,078	465.97	467.99	NHD	Continuous	Meters to nearest stream feature
Hydrologic variables						
Roughness	0.011-0.40	0.16	0.12	NLCD	Continuous	Manning's roughness based on land use/land cover classification
Imperviousness	0-100	29.78	32.64	NLCD	Continuous	Per cent impervious surface based on NLCD's land use/land cover classification
Hydraulic conductivity	0-11.7	0.44	0.82	USGS	Continuous	Centimetres/hour based on soil type
Socio-political variables						
Floodplain	—	—	—	HGAC	Discrete	In/out of the 100-year floodplain
Land use	—	—	—	HGAC	Discrete	12 land use classes as defined by H-GAC
Decade built	—	1979	—	HGAC	Discrete	Decade built
CRS participation	—	—	—	FEMA	Discrete	In/out of a CRS community

The variables were collected or derived from a variety of sources, including the: NED, National Elevation Dataset; NHD, National Hydrography Dataset; NLCD, National Land Cover Database; HCFCD, Harris County Flood Control District; HGAC, Houston-Galveston Area Council; FEMA, Federal Emergency Management Agency.

Dependent Variable: rescue requests were pulled on August 30, 2017 (n = 16,756) and subset for Harris County (n = 1,534) [8]

Test the model performance against three VGI datasets:

- 15% of rescue requests (n = 233)
- 311 calls (n = 1,328) [9]
- Flooded roads (n=1,000) [10]

Map flooding above a threshold (p(f) = 0.058) at which 95% of the rescue request test samples are predicted to be flooded

RESULTS

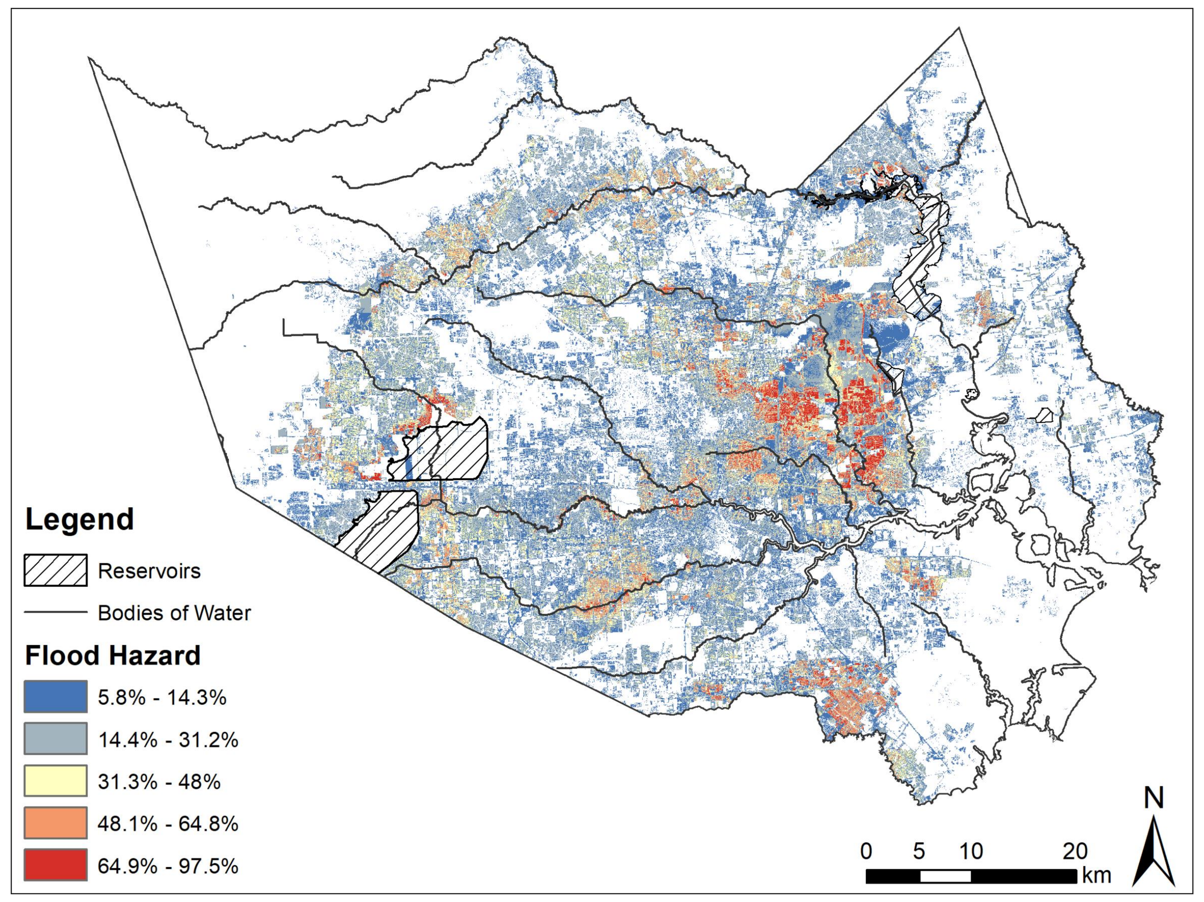


FIGURE 2 Probability of flooding in Harris County during Hurricane Harvey estimated using MaxEnt. Class intervals are created using standard deviations. The extent of flooding is estimated based on a threshold probability of 0.058

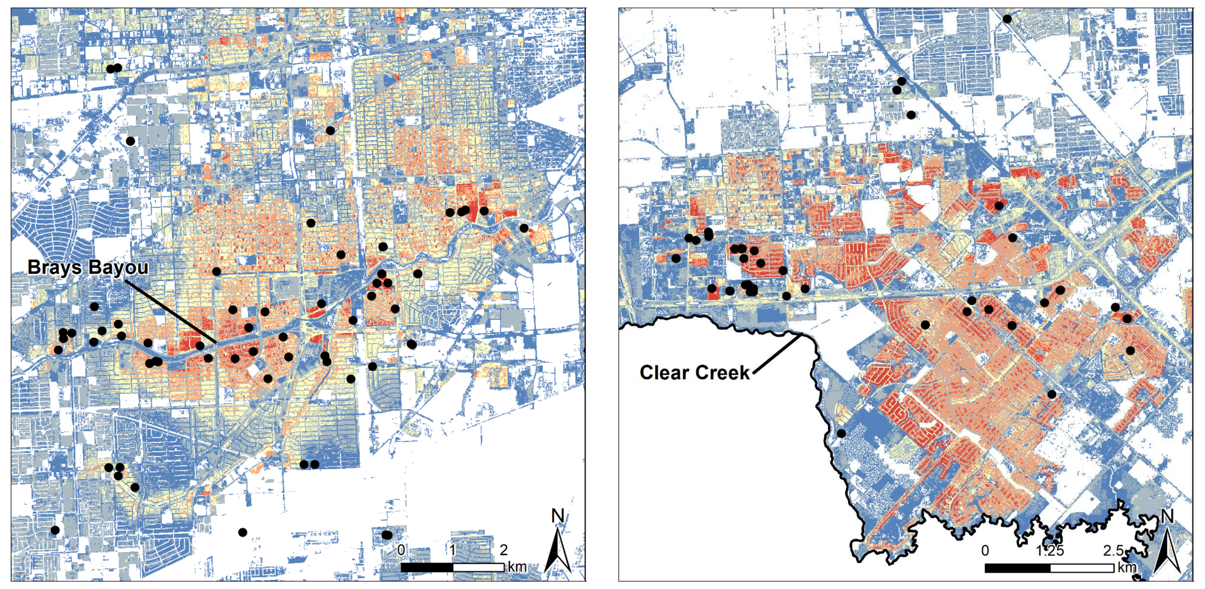
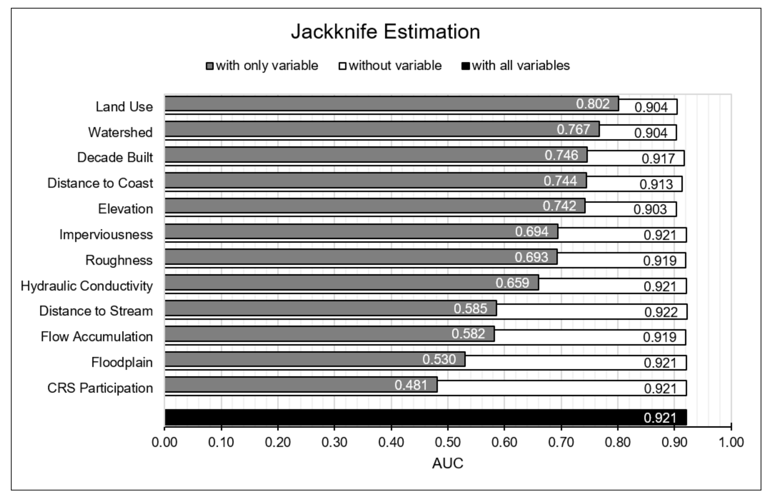


FIGURE 3 The jackknife estimation provides information about variable importance by testing the sensitivity of the MaxEnt model to the presence or absence of each variable. An AUC value is calculated for the entire model when including all variables (black) and can be compared against the AUC value calculated when including only one variable (grey) and when including all other variables (i.e., excluding that variable) (white) in the model. The longer the grey line, the more important that variable is for predicting the probability of a rescue request



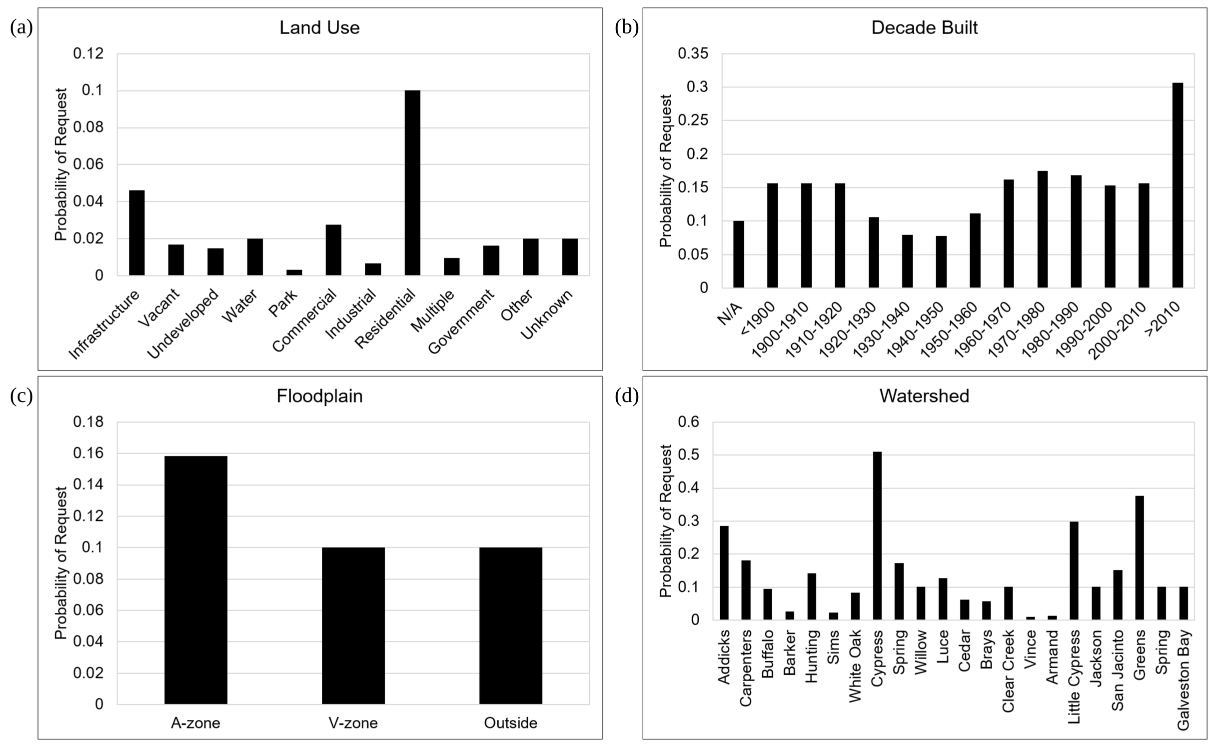


FIGURE 4 Model response curves for the categorical variables (a) land use, (b) decade built, (c) floodplain, and (d) watershed. The response curves indicate how predictive probabilities vary across a single variable when holding all others constant

CONCLUSIONS

- Model performs well against all three test datasets
  - Rescue requests (AUC 0.917)
  - 311 calls (AUC 0.929)
  - Flooded roadways (AUC 0.721)
- Individually, land use was shown to be the most important (AUC: 0.802), followed by watersheds (AUC: 0.767) and decade built (AUC: 0.746)
- Capable of rapidly identifying flooding in urban areas at a high spatial resolution
- Runs were performed in less than an hour (wall-clock time) on a personal computer
- Successful runs could start with as few as 5 requests
- Model fails to represent flooded areas with land uses that do not have rescue calls (e.g., parks and undeveloped land (areas behind Addicks and Barker dams)
- Effectiveness of the HDM for identifying flooding in rural areas is unknown

REFERENCES

1. van Oldenborgh, G. J., van der Wiel, K., Sebastian, A., Singh, R., Arrighi, J., Otto, F., … Cullen, H. (2017). Attribution of extreme rainfall from Hurricane Harvey. *Environmental Research Letters*, 12, 124009.
2. Blake ES, Zelinsky DA. (2018). National Hurricane Center Tropical Cyclone Report Hurricane Harvey (AL092017).
3. FEMA (2017). Historic disaster response to Hurricane Harvey in Texas. Retrieved from https://www.fema.gov/news-release/2017/09/22/historic-disaster-response-hurricane-harvey-texas.
4. Phillips, S. J., Anderson, R. P., & Schapire, R. E. (2006). Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, 190, 231–259.
5. Phillips, S. J., & Dudík, M. (2008). Modeling of species distributions with Maxent: New extensions and a comprehensive evaluation. *Ecography*, 31(2), 161–175.
6. Phillips, S.J., Dudík, M., & Schapire, R. E. (2016). Maxent software for modeling species niches and distributions (Version 3.3). Retrieved from http://biodiversityinformatics.amnh.org/open_source/maxent/.
7. Phillips, S. J., Anderson, R. P., Dudík, M., Schapire, R. E., & Blair, M. E. (2017). Opening the black box: An open-source release of Maxent. *Ecography*, 40(7), 887–893.
8. Crowdsource Rescue (2017). Hurricane Harvey Crowd Sourced Rescue Requests [Online]. Retrieved from: http://harveymaps.crowdrescuehq.org
9. City of Houston (2017). City of Houston 311 help & info [Online]. Retrieved from: http://www.houstontx.gov/311/.
10. Floodmap (2017). Crowdsourced Flooded Roadways [Online]. Retrieved from: https://floodmap.io.

ACKNOWLEDGEMENTS

This research was funded by the National Science Foundation PIRE Grant no. OISE-1545837

Citation: Mobley W†, Sebastian A†, Highfield W, Brody SD. Estimating flood extent during Hurricane Harvey using maximum entropy to build a hazard distribution model. *J Flood Risk Management*. 2019;12 (Suppl. 1):e12549.https://doi.org/10.1111/jfr3.12549

† these authors contributed equally to this work