Multiscale Data Integration and Modeling for Understanding Vulnerability During Disasters

Anna Brower[1], Julia Gohlke[2], Benjamin Zaitchik[3], and Samarth Swarup[1]

[1] University of Virginia, Charlottesville, VA 22903, USA
{aeb2de,swarup}@virginia.edu
[2] Virginia Tech, Blacksburg, VA 24061, USA
jgohlke@vt.edu
[3] Johns Hopkins University, Baltimore, MD 21218, USA
zaitchik@jhu.edu

Abstract. We describe the data integration and modeling framework in an ongoing project, which is creating an agent-based simulation of Hurricane Harvey. The goal of the project is to create a dynamic vulnerability index by augmenting the CDC Social Vulnerability Index, which would provide a better estimate of vulnerability and risk during disasters by taking human behaviors and mobility into account.

Keywords: Disaster modeling · Hurricane Harvey · Vulnerability.

Introduction

Vulnerability is broadly defined as the extent to which persons or things are likely to be affected by a hazard. The vulnerability of a human population group (in relation to other groups) is known as social vulnerability and is thought to be influenced by many factors, including demographics, neighborhood characteristics, the networks in which people are embedded, underlying health conditions, and more. In 2011, the US Centers for Disease Control and Prevention (CDC) introduced the Social Vulnerability Index (SVI) [4] to provide a measure of social vulnerability derived from census variables. The CDC SVI is increasingly being used in preparing for disasters and in planning relief and rescue efforts.

Health risk during a disaster is a spatiotemporally-dependent phenomenon in general that can be impacted by vulnerability [3]. People's risk varies as they move through areas of varying levels of hazards over the course of a day or week and social vulnerability may influence both the expected exposure to hazards as well as the outcome after exposure to a hazard. For example, exposure to extreme heat generally occurs in the middle of the day when temperatures are highest and when people may not be at home. Similarly, exposure to hazards like flash flooding are exacerbated when spending time in low-lying areas during periods of intense rainfall.

Estimating the spatiotemporal variability of risk, incorporating both social vulnerability and hazard exposure has always been limited by the lack of high resolution environmental and human mobility data. Lately, investigators have started turning to machine learning [10] and simulation-based methods [5, 9]. We consider the latter to be especially promising because they use multiple sources of data, integrated together to provide a model of human mobility patterns and consequent exposures to hazards. This offers the opportunity not just to estimate exposures, but also to evaluate the influence of vulnerability and ultimately potential interventions for mitigating risk.

In our current work, we are creating a detailed agent-based simulation model of Hurricane Harvey, in order to augment the CDC SVI and create a dynamic vulnerability index that takes into account human behaviors during disasters as well. In the following we briefly discuss some of the multi-scale data integration and modeling challenges in this ongoing work.

Hurricane Harvey

Hurricane Harvey was a Category 4 hurricane that was one of the most catastrophic and costliest hurricanes on record. It made landfall in August 2017, causing approximately $125 billion in damages and affecting 13 million people from Louisiana, Mississippi, Tennessee and Kentucky. With the hurricane making landfall three times in six days, the Houston area was flooded and thousands were forced to evacuate the area. In four days, areas around Texas received more than 40 inches of rain which caused flooding that peaked at 5 feet. With the damage inflicted by Hurricane Harvey, houses were left without power or were ruined beyond repair, forcing 30,000 residents of Texas to move into shelters [7].

Beyond the harm to infrastructure and housing, there were reports of damage to the mental and physical well-being of people affected by the storm. Sixteen percent of Texas Gulf Coast residents reported a worsening health condition or a new health condition. With an introduction to bacteria, dust and mold growth from damaged homes, there was an increase in respiratory problems. As well, following the hurricane there was an uptick in skin, eye, and ear infections after sewage-tainted water flooded the streets and waterways. Besides the damage to wastewater treatment plants, damage to industrial equipment exposed residents of the Houston-area to chemicals from pesticides, detergents and other common products contributing to conditions like nausea and eye irritation [2].

Multiscale Data Integration

Relevant data have been collected by many organizations and agencies, in multiple forms. Table 1 lists several of the data sets we have identified and are integrating into our agent-based simulation. These include data about environmental conditions (rainfall, flooding), damage to homes and infrastructure, data about the affected population, and about hazards such as Toxic Release Inventory (TRI) facilities.

To do data integration, we use two frames: a *person frame* and a *location frame*. Some data are naturally associated with people, such as demographics, activity patterns, health status, and behavior during the disaster. These are largely integrated from survey-type data, to construct a synthetic population [6].

Spatial data sets are at multiple resolutions and in multiple reference systems. The standard US Census hierarchy of regions is state – county – tract – blockgroup – census block [11]. In this reference system, smaller regions nest perfectly within the next larger level, e.g., blocks align perfectly with blockgroups, which align perfectly with census tracts, and so on. However, other reference geographies, such as neighborhoods, zip code regions, and grids can be of varying sizes and can overlap arbitrarily with the census regions.

Table 1. Data sets.

Data	**Description**	**Resolution**
Harvey Registry	Survey on impact from Harvey	Neighborhoods
Damage Assessment	Damage levels from inundation height	Latitude/Longitude
Hourly Rainfall	Hourly precipitation totals	Gridded region
Social Vulnerability Index	Quantification of vulnerabilities of regions	Tracts
Transportation Infrastructure	Condition of roads and bridges	Pathways of roads
TRI Facilities	Locations of toxic release inventory facilities	Latitude/Longitude
Inundation Raster	Presence of flooding	Latitude/Longitude
City of Houston Harvey Damage Assessment	Harvey affected property counts	Block group
Power Outages in CenterPoint Energy Service Area	Percentage of customers without power	Zip code
Individual Assistance Open Disaster Statistics	Data on registrations and Individuals and Households Program	County
American Community Survey	Demographic data	Block group
Spatial Hazard Events and Losses Database for US	Losses from Harvey	County
Harris County Flood Gauge Readings	Hourly rainfall and channel elevation readings from flood gauges	Latitude/Longitude

To integrate these, we take intersections of shapes of regions and are developing methods to assign variables from regions to resulting subregions. For example, hourly rainfall is available in a gridded reference frame through NLDAS [12]. To integrate this with the City of Houston Harvey damage assessment, which is given for each blockgroup, we take an intersection between blockgroup shapes and the NLDAS grid and assign damage estimates to each subregion of a blockgroup based on the rainfall in that subregion. A similar method has to be employed for integrating all the data sets with unaligned geographies.

This results in a data set with relatively small regions. To find the conditions at a given (latitude, longitude) location, we need to query this data set. To avoid having to do a large number of shape membership queries, we are developing an indexing scheme from grids to the smaller geographies that lie within them. A query point is first mapped to the grid cell that contains it, and from there to the regions that intersect with that grid cell. The goal of the data integration is to create a single data set that can be queried efficiently for all the variables at a given (latitude, longitude) location for modeling exposure to hazards, as described next.

Multiscale Modeling

Our agent-based simulations are not valid at the individual level, because synthetic individuals do not have a one-to-one mapping with the real inhabitants of the region. Synthetic populations are statistically accurate at the blockgroup level [1]. CDC SVI is constructed at the census tract level. During a hurricane or other disaster, people may engage in various behaviors, such as evacuation, sheltering-in-place, shopping for supplies, picking up family members, and more [8]. These result in complex spatiotemporal trajectories, crossing multiple tracts and blockgroups, with varying exposure to hazards.

Thus the overall agent-based simulation involves integrating data at multiple scales, as described in the previous section, modeling individual movements through these regions, and integrating exposures back to the census tract level. The effects of exposure eventually show up in syndromic surveillance data sets, when people report symptoms such as infections, rashes, nausea, eye irritation, etc. [2]. We plan to use these data sets to validate our model.

Conclusion

An improved understanding of the spatiotemporal variability of population vulnerability will lead to better preparedness and planning for future disasters. Risk is contingent upon local circumstances, such as the specific nature of the disaster, the topography, the built environment, and population demographics, and spatiotemporal patterns of vulnerability will vary across regions and types of disaster. Our goal is to demonstrate that with the right tools for data collection, analysis, and simulation in place, we can rapidly generate models and possibly forecasts of vulnerability in advance of the next major hurricane to help mitigate its effects.

References

1. Beckman, R.J., Baggerly, K.A., McKay, M.D.: Creating synthetic base-line populations. Transportation Research A – Policy and Practice **30**, 415–429 (1996)
2. Constible, J.: The emerging public health consequences of Hurricane Harvey (August 2018), https://www.nrdc.org/experts/juanita-constible/emerging-public-health-consequences-hurricane-harvey
3. Cutter, S.L., Boruff, B.J., Shirley, W.L.: Social vulnerability to environmental hazards. Social Science Quarterly **84**(2), 242–261 (2003)
4. Flanagan, B.E., Gregory, E.W., Hallisey, E.J., Heitgerd, J.L., Lewis, B.: A social vulnerability index for disaster management. Journal of Homeland Security and Emergency Management **8**(1), Article 3 (2011)
5. Karner, A., Hondula, D.M., Vanos, J.K.: Heat exposure during non-motorized travel: Implications for transportation policy under climate change. Journal of Transport & Health **2**, 451–459 (2015)
6. Mortveit, H.S., Adiga, A., Barrett, C.L., Chen, J., Chungbaek, Y., Eubank, S., Kuhlman, C.J., Lewis, B., Swarup, S., Venkatramanan, S., Wilson, A., Xie, D., Marathe, M.V.: Synthetic populations and interaction networks for the U.S. Tech. rep. (2020), nSSAC Technical Report: #2019-025
7. NOAA National Centers for Environmental Information: U.S. billion-dollar weather and climate disasters (2020), https://www.ncdc.noaa.gov/billions/events
8. Provitolo, D., Dubos-Paillard, E., Müller, J.P.: Emergent human behaviour during a disaster: Thematic *Versus* complex systems approaches. In: Proceedings of EPNACS 2011 within ECCS'11. Vienna, Austria (September 15 2011)
9. Swarup, S., Gohlke, J.M., Bohland, J.R.: A microsimulation model of population heat exposure. In: The 2nd International Workshop on Agent-based Modeling of Urban Systems (ABMUS). São Paulo, Brazil (2017)
10. Terti, G., Ruin, I., Gourley, J.J., Kirstetter, P., Flamig, Z., Blanchet, J., Arthur, A., Anquetin, S.: Toward probabilistic prediction of flash flood human impacts. Risk Analysis **39**(1), 140–161 (2019). https://doi.org/10.1111/risa.12921
11. United States Census Bureau: Geography: Hierarchy diagrams (Last revised: Dec 5, 2017), https://www.census.gov/programs-surveys/geography/guidance/hierarchy.html, accessed on Feb 28, 2020
12. Xia, Y., Mitchell, K., Ek, M., Sheffield, J., Cosgrove, B., Wood, E., Luo, L., Alonge, C., Wei, H., Meng, J., Livneh, B., Lettenmaier, D., Koren, V., Duan, Q., Mo, K., Fan, Y., Mocko, D.: Continental-scale water and energy flux analysis and validation for the North American Land Data Assimilation System project phase 2 (NLDAS-2): 1. intercomparison and application of model products. Journal of Geophysical Research: Atmospheres **117**(D3) (2012). https://doi.org/10.1029/2011JD016048